DOCUMENT RESUME

ED 338 691	TM 017 520

AUTHOR	Dorr-Bremme, Donald W.; Herman, Joan L.
TITLE	Assessing Student Achievement: A Profile of Classroom Practices. CSE Monograph Series in Evaluation 11.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	86
NOTE	132p.
PUB TYPE	Statistical Data (110) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Academic Achievement; Classroom Techniques; Educational Assessment; *Elementary School Teachers; Elementary Secondary Education; Evaluation Utilization; Instructional Leadership; National Surveys; *Principals; *Secondary School Teachers; *Student Evaluation; Tables (Data); Teacher Made Tests; Testing Programs; *Test Use
IDENTIFIERS	Testing Effects; *Test Use in Schools Study

ABSTRACT

The Center for Evaluation Study at the University of California (Los Angeles) conducted the Test Use in Schools study to map the topography of basic-skills achievement testing and achievement test use in public schools across the United States. The survey addressed a nationwide sample of principals and teachers through a successive, random-selection procedure, and focused on determining how leadership activities impact test use. Responses were obtained from 220 principals, 475 elementary school teachers, and 363 high school teachers in 91 of the 114 districts sampled. Return rates were about 50% for high school teachers, about 60% for principals, and about 60% for elementary school teachers. Examination of the survey responses confirms that there are two tiers of student achievement assessment in the country. One tier is internal or local to the schools; it is "owned" and usually produced by the teachers themselves as tests and other assessments of achievement in the classroom by the individual student. The other tier of assessment is external to the school; it is mandated by district, state, or federal program requirements. Both tiers are underutilized; neither fulfills its potential for helping students. Case studies and research-based models are included to illustrate how schools and districts can integrate these tiers to plan for instructional improvement. There are 31 tables of study data and 6 figures illustrating the concepts discussed. A 50-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED338691

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. C. BEER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CSE
MONOGRAPH
SERIES
IN
EVALUATION

11

ASSESSING STUDENT ACHIEVEMENT: A PROFILE OF CLASSROOM PRACTICES

Authored by
Donald W. Dorr-Bremme
and
Joan L. Herman

TM017520

CENTER FOR THE STUDY OF EVALUATION
UNIVERSITY OF CALIFORNIA • LOS ANGELES
1986

2 BEST COPY AVAILABLE

CSE MONOGRAPH SERIES IN EVALUATION

SERIES EDITOR

Eva L. Baker

Center for the Study of Evaluation
UCLA Graduate School of Education
University of California, Los Angeles
Los Angeles, California 90024

ASSESSING STUDENT ACHIEVEMENT: A PROFILE OF CLASSROOM PRACTICES

Authored by
Donald W. Dorr-Bremme
and
Joan L. Herman

4

CSE MONOGRAPH SERIES IN EVALUATION

NUMBER

1. Domain-Referenced Curriculum Evaluation: A Technical Handbook and a Case Study from the MINNEMAST Project
   Wells Hively, Graham Maxwell, George Rabehl, Donald Sension, and Stephen Lundin
2. National Priorities for Elementary Education
   Ralph Hoepfner, Paul A. Bradley, and William J. Doherty
3. Problems in Criterion-Referenced Measurement
   Chester W. Harris, Marvin C. Alkin, and W. James Popham (Editors)
4. Evaluation and Decision Making: The Title VII Experience
   Marvin C. Alkin, Jacqueline Kosecoff, Carol Fitz-Gibbon, and Richard Seligman
5. Evaluation Study of the California State Preschool Program
   Ralph Hoepfner and Arlene Fink
6. Achievement Test Items — Methods of Study
   Chester W. Harris, Andrea Pastorok Pearlman, and Rand R. Wilcox
7. The Logic of Evaluative Argument
   Ernest R. House
8. Toward A Methodology of Naturalistic Inquiry in Educational Evaluation
   Egon G. Guba
9. Values, Inquiry, and Education
   Hendrik D. Gideonse, Robert Koff, and Joseph J. Schwab (Editors)
10. Evaluation in School Districts: Organizational Perspectives
    Adrianne Bank and Richard C. Williams with James Burry (Editors)
11. Assessing Student Achievement: A Profile of Classroom Practices
    Donald W. Dorr-Bremme and Joan L. Herman

This Project was supported in whole or in part by the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

TABLE OF CONTENTS

Acknowledgements vii

Chapter 1 Introduction 1

Chapter 2 Assessing Student Achievement: The Frequency of Testing and the Time it Takes 15

Chapter 3 Using Assessment Results 29

Chapter 4 Administrative Leadership: Monitoring and Supporting Assessment 45

Chapter 5 Principals' and Teachers' Perceptions and Beliefs About Testing 57

Chapter 6 The School Context and Classroom Testing Practices 79

Chapter 7 Summary and Implications: Issues for State and National Policy Makers 93

Chapter 8 Directions For Policy and Practice at the Local Level: Linking Testing With Instructional Planning and Improvement 109

ACKNOWLEDGEMENTS

This monograph is the result of a five-year project conducted by the Center for the Study of Evaluation with funding from the National Institute of Education (NIE). The project would not have been possible without the great contributions of many individuals:

First and foremost, our thanks to Eva Baker, Director of the Center, who made the project possible. Dr. Baker suggested the need for the project, helped guide its initial design, and contributed her ideas, enthusiasm, and support throughout the entire project.

Our thanks also to the many CSE staff members who participated with us: Jennie Yeh, who co-directed the study during its first critical year; Bruce Choppin and William Doherty, who helped to direct its data analysis; James Burry, who contributed significantly to every phase of the project; Linda Polin, Charlotte Lazar-Morrison and Raymond Moy who ably conducted the literature review; Beverly Cabello and Liza Daniels, who provided critical assistance in the design and implementation of the field study and testing costs; James Lehman and Chih-Ping Chou who competently organized the data and responded to our innumerable requests for new analyses; and to Aeri Lee and Katharine Fry who patiently and ably managed the production of this manuscript.

We are grateful also to Lawrence Rudner, our NIE monitor, who continuously encouraged our work and to the Institute itself for the support it provided.

Finally and especially, we would like to thank the district administrators, school principals, teachers and others in the schools who cooperated in the study. We greatly appreciate the time they donated to us, filling out our questionnaires, responding to our questions, and sharing their perceptions with us.

To all who participated with us and made our work possible, we extend our sincere appreciation for their important contributions.

Donald W. Dorr-Bremme
Joan L. Herman

CHAPTER 1

INTRODUCTION

Fueled by school board accountability concerns, minimum competency mandates, evaluation requirements for federal, state and local programs, and the growth of curriculum-embedded and continuum-based assessment systems, achievement testing in American schools has become both an enterprise of significant scope and visibility and the subject of considerable public discussion and debate. Critics have attacked the arbitrariness of current testing practices (Baker, 1978), have expressed concerns about their validity and bias (Perrone, 1978), have accused testing of narrowing the curriculum and have questioned the value of traditional testing amidst changing functions of education (Tyler, 1977). The quality of available tests continues to be controversial (CSE, 1979; The Huron Institute, 1978), at least one major teachers' organization called for a moratorium on the use of standardized tests, and vigorous legal battles have been launched.

Responding to these various challenges, adovcates of testing have reaffirmed its importance and reasserted the variety of purposes that current tests can and do serve. Supporters have maintained, for example, that testing promotes accountability, facilitates more accurate placement and selection decisions, and yields information useful for curricular and instructional improvement.

The testing controversy rages on while the nation's considerable investment in achievement testing continues. Although the stakes in the debate are high, public policy in this arena has been formulated without the benefit of basic information about the nature of testing as it actually occurs and is used in schools. How much testing really goes on? How are test results used? What functions do tests serve for teachers and principals? What are the effects on schools of various local, state and federal manadates? These and similar questions have gone largely unaddressed. A few studies have indicated teachers' reservations about the limited use of one type of achievement measure — the norm-referenced standardized test (Airasian, 1979; Boyd et al, 1975; Goslin, 1965; Goslin, Epstein, & Hallock, 1965; Resnick, 1981; Salmon-Cox, 1981; Stetz & Beck, 1979). Beyond this, however, the landscape of testing practices and test use in American schools remains largely unexplored.

In this context, the UCLA Center for the Study of Evaluation's (CSE) three-year study provides educational policy-makers with basic, new information on classroom achievement testing across the United States. Conducted from 1979 through 1983, CSE's research was designed to take a comprehensive picture of national testing practices. It investigated a wide range of types of formal assessment measures (e.g., commercially produced norm- and criterion-referenced tests and curriculum embedded measures, tests of minimum competency and functional literacy; district-, school-, and teacher-developed tests) as well as some less formal means for gauging student progress and achievement (teachers' observations of and interactions with learners). Within this broad range, inquiry focused on achievement testing practices in reading/English and in mathematics, basic skills areas which are the subject of continuing public concern. Teachers and principals at both elementary and secondary grade levels served as primary subjects for the nationwide survey, addressing those grade levels which had been identified in prior research as important transition points and the targets of frequent testing. The research commenced with an extensive literature review and exploratory fieldwork in three school districts across the country to identify relevant contextual variables and to deepen our understanding of teachers' and principals' orientations. Case study inquiry following the survey explored in greater detail issues associated with the costs of testing.

Policy Orientation: Questions and Issues of Interest

As the discussion above suggests, educational achievement testing is a pervasive enterprise, one which recurrently affects the lives of all students. It is an enterprise which is rapidly changing, diversifying and expanding. And it is an enterprise in which hundreds of millions of dollars in public monies are expended annually. It is not surprising, then, that it generates a broad range of questions and issues for policymakers to address. The CSE study examined a number of these:

Competency testing. Across the nation, more than 40 states have now mandated tests of minimum competency for school children. Some states require such tests for promotion and graduation; others for checking students' basic educational needs at milestones in their school careers. Decisionmakers at all levels need to know how these testing programs are influencing students' educational experiences and life chances. What are the impacts of different kinds of minimum competency programs? Have they affected curriculum and instruction? Have they wrought changes in the other ways districts and schools measure students' progress?

Testing for federal and state program evalution. Federal and state categorical programs, meanwhile, continue to include evaluation requirements. Testing student achievement remains a primary way of meeting those requirements. Program administrators and technical assistance personnel in both funding agencies and participating districts, along with legislators and their advisors, need cost benefit information on testing in this context. Can it and does it serve purposes beyond accountability and compliance? How does testing for federal and state program evaluation affect the instructional time of participating students? How does it influence the distribution of instructional staff members' energies and efforts?

District continuum testing. Simultaneously to the above activities, many school districts are expanding their own testing programs. And increasingly these district tests monitor students' progress along district-mandated sequences (or continua) of skills or objectives. From district to district, however, teachers may differ in their willingness to administer such tests and to utilize the results. Under what conditions, then, are test accompanying skills continua most likely to be administered and used in instructing students? What qualities should the tests have to be maximally useful? How can they be effectively integrated with other assessment activities? District administrators require information to resolve these issues.

Teacher-constructed tests and other assessment techniques. Teachers themselves seem to spend significant amounts of their assessment time in administering tests and quizzes that they construct. They also seem to devote considerable attention, especially in the elementary grades, to commerically produced tests that come with curriculum materials. What are the qualities in these kinds of tests that make them attractive and useful?

Defining the Research Problem

Given the vast array of policy issues and information needs surrounding educational testing, how should a national student survey be focused? CSE's Test Use Survey was guided by two interrelated concepts:

— the concept of the teacher as practical reasoner and decisionmaker;

— the concept of testing as an intervention

The teacher as practical reasoner and decisionmaker. The view of teachers as practical reasoners and decisionmakers emerges from theory and research from the branch of sociology known as ethnomethodology (Cicourel, 1974; Garfinkel, 1967; Cicourel, & Kitsuse, 1963; Mehan &

Wood, 1975; Weider, 1973; Wood, 1968). According to this view, as practical reasoners and practical decisionmakers, members of social units:

- — Orient their activities to the practical tasks they must accomplish in their everyday routines and do so in light of the practical contingencies and exigencies they face;
- — Carry out their activities based on their "background understandings" of a "world known in common and taken for granted" (Schutz, 1962). That world is validated and supported daily through members' collective activities. Members act as "naive phenomenologists," taking things as they seem to be until unfolding experience proves them otherwise. Thus they sustain their orientations to their practical tasks and circumstances.

Data from the Test Use in School Study's planning-stage fieldwork efforts support such a view. That teachers do orient their efforts to the practical tasks that are demonstrably central in their everyday professional lives and do orient to the practical exigencies they face was recurrently documented. Teachers, for example, reported their uses of test results as serving most heavily the functions that are central to their routine teaching responsibilities: deciding what to teach and how to teach it to students of different achievement levels; keeping track of how students are progressing; and evaluating and grading students on their performance (Dorr-Bremme, 1983). Further, the means of assessment that teachers reported using most often and in the greatest variety of ways were those which facilitate the accomplishment of their practical activities and respond to the practical exigencies they face.

A variety of routine tasks constitute the world of teaching as practiced. Teachers must accomplish these tasks in a context characterized by recurrent time limits, others' demands for high performance and accountability, and their own concerns with providing effective and appropriate instruction. These features of the teaching world impinge upon teachers' testing practices and test use. Thus, it appears that their reasoning and decision-making about asssessment and its uses are structured by and oriented to their practical circumstances.

Testing as an intervention. A second concept framing the Test Use in Schools Survey was the concept of testing as an intervention. From this perspective, required or recommended tests, by virtue of their very presence in schools can impact educational practices. They can, in fact, function as change agents. Supporting this point of view, planning stage research indicated that:

1. Mandated tests can add new standards of accountability to those that teachers must attend to in their everyday routines. Reasoning

practically, teachers may feel responsible for adjusting their instructional emphases and techniques to match the skills and information students must master to do well on required tests. For example, minimum competency tests, particularly those required for graduation, seem especially likely to re-orient teachers' practical reasoning and instructional planning and induce them, individually and schoolwide, to alter curriculum and teaching methods.

2. Mandated tests can change the practical circumstances under which teaching and learning must be accomplished. Respondents in the exploratory field research, for instance, cited a number of unintended, largely negative, effects of testing programs, e.g, reduction in time for teaching. Where consequences of this type occur, they alter the practical contingencies that teachers face in accomplishing their routine activities. As they do, they may occasion broader changes in instructional practices, curriculum, and perhaps in students' learning as well.

3. Mandated tests, where they respond to teachers practical exigencies, can provide new ways to accomplish routine tasks and can signal new approaches to instructional practice. Fieldwork in two districts, for example, illustrated the ways in which a district continuum test can respond to teachers' assessment needs and facilitate more individualized instructional approaches. Under such circumstances, testing programs of particular kinds can serve as agents for educational change.

Framework for the National Survey

The two related concepts of the teacher as a practical reasoner and testing as an intervention provided a useful organizing framework for the national survey of assessment practices and uses schools and classrooms. In addition to informing the selection of domains to be examined in survey questionnaires, this framework indicated some interesting relationships to be explored. These domains and hypothetical relationships are displayed in Figure 1. (Notice that not all the relationships portrayed there were examined in the national survey.)

Federal/state/local testing requirements. Attention to such requirements responds to the concept of testing as an intervention. As depicted, testing requirements influence the distribution and frequency of types of testing at local sites, and thus bear upon patterns of test use. (That is, districts may introduce innovative tests that teachers use heavily to replace self-constructed tests, etc. Federal and state evaluation requirements may encourage consolidation of assessment activities and use of extant tests for "new" purposes, or they may simply introduce additional testing at local sites.) Following the chain of posited relationships further, testing interventions such as minimum competency programs may impact on the organization of curriculum and instruction (as described above).

Given that types of assessment seem to impact on one another and given the seeming importance of minimum competency testing as an agent of change, districts were sampled on presence/absence of statewide assessment and on various conditions of minimum competency testing. Data on the federal-, state-, and district-initiated testing in sampled districts and schools were elicited in brief, initial, district-contact phone interviews with district testing officers and through principal questionnaires.

Federal/state/local programs. The presence/absence of particular federal and state categorical programs, and local educational programs as well, is assumed to influence how curriculum and instruction are organized in schools and, in turn, the routine tasks of local-site practitioners. (For instance, Title I and Title VII programs and programs developed in response to Public Law 94-142 occasion referral, placement, and diagnostic decisions.) The testing that occurs and the test scores that are used follow from needs inherent in these routine tasks. The study was not explicitly interested in studying how federal, state, or local programs impact on the organization of curriculum and instruction locally (dotted line, arrows). It was only interested in the presence/absence of the instructional alternatives such programs provide. Thus, only information on district and school participation in major, instruction-related federal and state programs, e.g., Title I, (Chapter 2) was gathered.

Organization of curriculum and instruction. The organization of curriculum and instruction constitutes a main influence on the nature of teachers' routine, practical activities and decisions. If students are grouped by reading level or set to work in individualized, self-paced learning programs, the teachers need to make placement decisions. If a continuum of objectives or "management system" is established then teachers must monitor learners' progress through that continuum. If team teaching is practical or aides are available for instructing students, students must be distributed to the instructional alternatives afforded by extra personnel (Yeh, 1978; Yeh, 1980). In summary, it was hypothesized that a greater variety and number of available instructional alternatives in the classroom and school would increase the routine tasks and decisions that require assessment information, and so influence both the patterns of testing that occur locally and the ways test scores are used locally.

Data on the organization of curriculum and instruction were gathered primarily on teacher questionnaires: e.g., the presence/absence of aides and team teaching, the ways teachers distribute students for instruction within the class, presence and type of instructional support services beyond the classroom. Information on the latter was also elicited from principals.

Types of students served. The nature of practitioners' routine, practical activities and decisions was assumed to vary with the types of students enrolled in the school and assigned to a teacher's classroom. Students whose first language is not English, who are members of socioeconomically depressed and/or culturally different populations, whose rate of achievement is unusually rapid, and so on, present teachers with different kinds of instructional challenges and decisions. Thus, the types of testing given locally and the uses of test results are likely to vary with the demographic or achievement characteristics of children in the school and classroom.

Breakdowns of sampled schools' enrollments by socioeconomic status (as indicated by percent receiving Aid to Families with Dependent Children, percent receiving free lunch, and similar indices) and ethnic identity were elicited from principals. Principals were also asked to provide contextual information on the rate of transience in school enrollment year-to-year and on recent general enrollment trends.

Teachers' perceptions of the utility of tests and types of tests. As teachers go about the accomplishment of their practical tasks and decisions, the instances in which they refer to test scores and the ways in which they "count" or "weigh" test scores are assumed to vary with their perceptions (opinions, values, understandings) of tests and types of tests (See Lazar-Morrison et al., 1980; Yeh, 1980).

Survey instruments for teacher respondents gathered data on teachers' perceptions and beliefs about testing particular types of tests and testing in general.

Teachers' experience and training. As they go about making sense of particular tests' strengths and weaknesses, appropriate uses, and the like, teachers (the model assumes) will draw upon their formal educational and practical experiences with respect to testing. Thus, their training and experience are likely to bear ultimately on their practical decisions about which types of test scores to use and how to use them. Teacher questionnaires asked respondents to report succinctly on the number of years they have been teaching and the number of years they have been teaching in their present school. (The latter was assumed to index teachers' familiarity with existing local assessment programs and practices, socialization to local norms and values, etc.) Information on teachers' educational background knowledge and in-service training experience also was elicited.

District and local site leadership action. It was assumed that innovative district and school leadership can provide in-service training experiences that change teachers' perceptions of the utility of particular tests and types of tests, thus influencing teachers' practical test-use decisions. District and school leaders can also, it was posited, act to generate tests, testing programs, and testing practices that facilitate teachers' accomplishment of their routine tasks under the practical exigencies of their environments (See Dorr-Bremme, 1983). Finally, district and school leaders may act to require that teachers use certain test scores for particular purposes.

The study was not explicitly interested in how types of leadership action impact on types of in-service training in testing (dotted lines, arrows). The study was interested, however, in how leadership activities of particular kinds impact on test use (solid line, arrows). Data on district-wide leadership action were collected in initial-contact phone interviews with district testing officials and on principal questionnaires. Information on school-site leadership was gathered from teacher questionnaires.

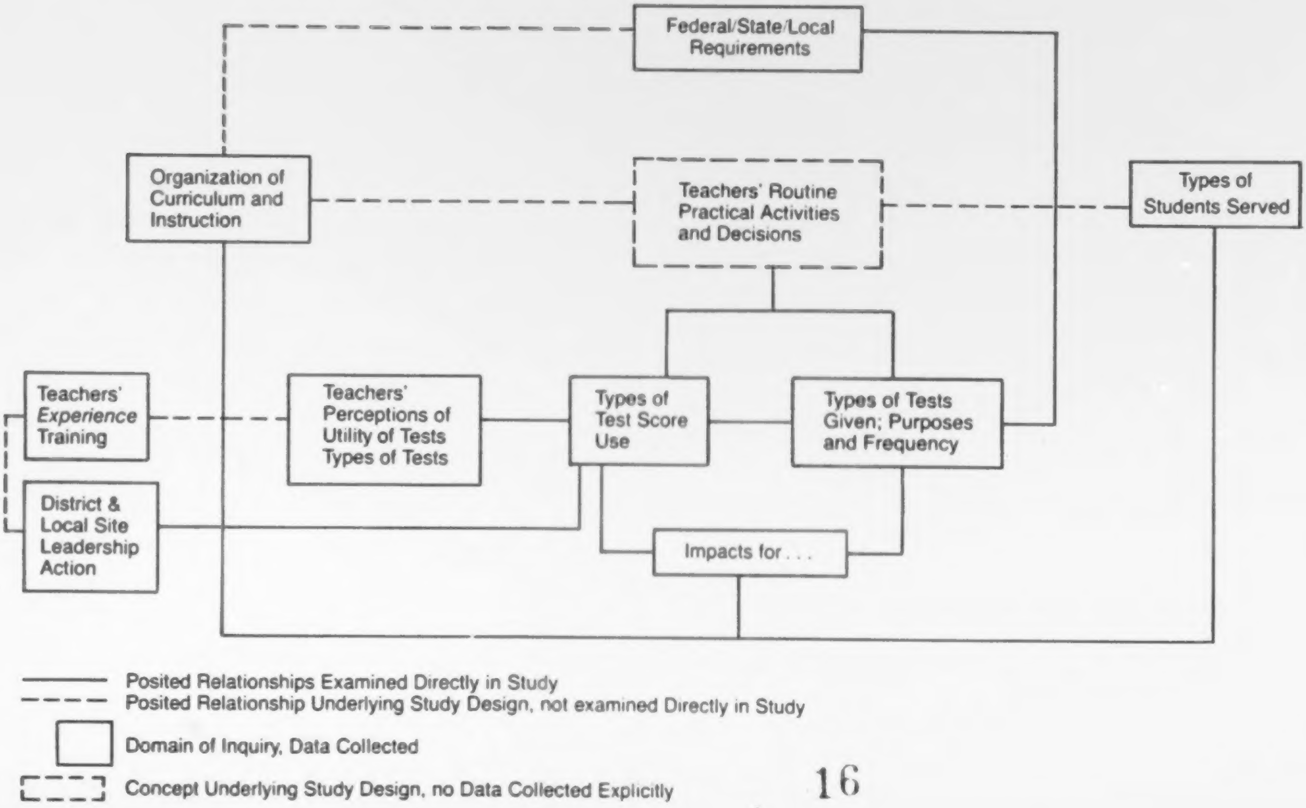
Types of tests given: purposes and frequency. Describing the types of tests given at local school sites was a central goal of the study. So too was identifying the factors that influence the purposes for tests and the frequency with which they are given; hence the inclusion of the domains discussed in the foregoing paragraphs.

The model assumed that the types of test given locally, and the purposes for and frequency with which they are given, will influence local types of test-score use. This assumption was made for more than the obvious reason, that the giving of a type of test makes its scores available. It was also posited that the presence/absence of one type of test may influence the use of scores from another type. The giving of minimum competency tests as a requirement for graduation, for instance, may encourage teachers to use the results of other kinds of tests to measure students' progress toward attainment of the minimum competencies. (This phenomenon was observed in a junior high school visited during exploratory field work.) Similarly, the absence of particular types of testing in a local setting may co-occur with more diverse uses of the results of tests that are given there.

Data on the types of tests given, and on the purposes for and frequency with which each is administered, were elicited from both teachers and principals, assuring a comprehensive picture of the pattern of testing in each school and classroom sampled.

Figure 1

Conceptual Model Guiding — Test Use Survey Inquiry


Federal/State/Local Requirements
Organization of Curriculum and Instruction
Teachers' Routine Practical Activities and Decisions
Types of Students Served
Teachers' *Experience* Training
Teachers' Perceptions of Utility of Tests Types of Tests
Types of Test Score Use
Types of Tests Given; Purposes and Frequency
District & Local Site Leadership Action
Impacts for . . .
Posited Relationships Examined Directly in Study
Posited Relationship Underlying Study Design, not examined Directly in Study
Domain of Inquiry, Data Collected
Concept Underlying Study Design, no Data Collected Explicitly

Types of test score use. Describing how scores from particular types of tests are actually used was another primary goal of the research. And identifying the factors that influence type-of-test-score/type-of-test-use relationships was yet another.

Information on how scores from particular kinds of tests are used in classrooms was elicited on teacher questionnaires. Data on other school-wide uses of test scores was gathered on principal questionnaires.

Impacts. As Figure 1 shows and as earlier discussion has explained, it was assumed that testing can have influence within schools in two ways. First, testing can have influence through practitioners' use of test scores in decision making. For example, curriculum program and/or instructional strategies might be changed in response to a program evaluation including test scores as measures of program effectiveness. Test scores might influence student placement decisions. Second, tests can impact on curriculum and instruction by virtue of their very presence as required or recommended. In the study's conceptual framework, then, both "types of test score use" and "types of tests given" are assumed to have potential impact.

The conceptual model also calls attention to the study's interest in the impacts of particular types of testing and test-score use for learners in general and for particular types of learners (referenced as "types of students served"). The model also indicates the interest of the research in impacts of particular types of testing and test-score use on curriculum and instructional activities. These potential impacts were discernible in the research through:

(1) Questionnaire items that investigate the ways in which test scores are used.
(2) Questionnaire items that asked about respondents' perceptions of the impacts of particular types of testing on their students, classrooms, and schools.
(3) Data analyses that examined relationships between types of students served (e.g., by socioeconomic condition and amount of testing, types of tests given, and patterns of test score use).

The Survey Sample*

The survey addressed a nation-wide sample of principals and teachers drawn through a successive, random-selection procedure. Given the study's intent to provide a comprehensive picture of current testing practices, sampling procedures were devised to yield a nationally representative sample of respondents. Stratifying variables reflected this concern for representativeness, as well as the need for variables whose values were easily attainable; these included geographic region of the country, district size, urban-suburban-rural locale, socioeconomic status, and minimum competency testing policy. The latter two variables also reflect the study's interest in clarifying policy issues, though the number of policy-relevant sampling variables which could be included in sampling was severely limited by available information. While it might have been interesting to stratify the sample based on district leadership or types of district-required tests, for example, no prior information existed which would permit selections based on these variables.

Respondent sampling proceeded as follows. First, a nationally representative probability sample of 114 school districts was drawn. (A lattice sampling technique was used to select cells from the matrix defined by the five stratifying variables. Then random sampling was done to select within cells.) Next, from within these districts, size permitting, two elementary schools and two high schools were randomly selected using a procedure that facilitated (where possible) inclusion of schools at levels serving both higher- and lower-income populations. Finally, in each of these schools, principals received directions for randomly drawing four teachers for inclusion in the study. Directions for elementary principals guided the random selection of two fourth-grade and two sixth-grade teachers; those for high school principals directed the random selection of two teachers of tenth-grade English and two of tenth-grade mathematics.

The principal and each of the four participating teachers at each school received questionnaires that elicited detailed information on their individual and school testing practices, as well as related contextual and attitudinal data.

Return rates. Returns were obtained from 220 principals, 475 elementary-school teachers, and 363 high-school teachers in 91 of the 114 districts sampled. Return rates from all principals and from teachers at the elementary level were approximately 60%. About 50% of the high school teachers in the sample responded. To correct for differential return rates by sampling cell, and to approximate a nationally representative distribution of respondents, weightings were applied in all descriptive analyses. The results reported in the following chapters, therefore, represent weighted estimates of national testing practices, test use patterns, and principal and teacher perceptions and beliefs on testing-related issues.

---

*A more detailed description of the sampling procedures is available in Burry et al., 1982

What was the nature of the selected schools, their teachers and classrooms? In order to provide context for understanding the results presented in later chapters, the remainder of this section describes the characteristics of the school environment in which the respondents operate and then the teachers themselves.

The average elementary school in the sample served a total enrollment of 528, comprised of a majority Caucasian but ethnically mixed student population. While the typical school community was economically heterogeneous, a significant minority of students received federal aid and/or qualified for free school lunch benefits. Transiency and absence rates were relatively modest, 16 and 6 percent respectively. A majority of the schools (60%) operated a school improvement program, and student achievement testing was typically included and required in such programs. Over one-half of the schools operated under minimum competency testing requirements; while within these schools most students passed such required tests on the first try, a sizeable number of students (20%) typically experienced failure. (See Table 1)

Secondary school enrollments, as would be expected, were substantially higher, with a mean of 1439. While other characteristics were quite similar to those at elementary school level, students in the average high school in the sample appeared slightly more economically advantaged and less transient.

The average teacher within the schools described above had approximately twelve years of teaching experience, almost ten of which were in their current district. (The results are presented in Table 2.) In terms of their education the respondents were almost evenly split between those holding a Bachelors degree and those holding a Masters degree, with less than 1% holding a doctorate. Further, they tended to average some 24 to 25 college units beyond their highest degree. The picture of the teachers then, is one of experienced, educationally qualified professionals who have continued to pursue education. It is interesting to note how similar the characteristics were across the elementary and secondary levels. At both levels, however, these characteristics appeared unrelated to testing practices.

The routine of the classrooms these teachers taught in is also described in the results found in Table 2. The results indicate that teachers had in their classrooms approximately 27 students at the elementary level and 26 at the secondary level. At the elementary level, they provided over 6.5 hours of reading instruction per week and about 5 hours of mathematics instruction. The results at the secondary level were similar for mathematics, i.e., about 5.5 hours of instruction per week. However, fewer hours of English instruction occurred at the secondary level (approximately 5.5 hours) than reading instruction at the elementary level, reflecting both the greater emphasis on reading earlier in a student's career and the broadening of the curriculum as a student progresses through higher grade levels, as well as standard class periods at the secondary level. It will be useful to compare these average hours of weekly instruction with the amount of time devoted to testing. This is done in the next chapter, where the frequency of testing and the time it takes are described.

Table 1. School Characteristics

	Elementary		Secondary	
	Mean	**S.D.**	**Mean**	**S.D.**
Total Enrollment	528	(235)	1439	(696.3)
School Ethnicity				
Black	15.0%	(25.8)	15.0%	(25.5)
Hispanic	8.1%	(21.2)	6.8%	(18.4)
Asian	2.1%	(9.2)	0.7%	(1.2)
Native American	5.5%	(20.4)	0.4%	(2.1)
Caucasian (Euro-American)	70.6%	(35.8)	76.2%	(31.0)
Other	1.2%	(9.9)	0.7%	(5.7)
Socio-Economic Status				
Low income (<$8,000)	29.0%	(26.2)	22.4%	(20.2)
Middle income	50.6%	(23.4)	56.7%	(19.3)
High income (>$25,000)	20.5%	(21.7)	21.8%	(17.6)
% of students receiving AFDC or free lunch	31.0%	(26.2)	23.2%	(22.8)
Transiency Rate	15.5%	(13.7)	10.4%	(7.8)
Absentee Rate	6.0%	(9.4)	7.4%	(3.7)
School Improvement Program				
% Participating	59.7%	—	63.0%	—
% Requiring Testing	76.3%	—	65.7%	—
Minimum Competency Testing				
Required	53.3%	—	50.0%	—
% Students passing first time	80.0%	(23.0)	76.1%	(22.6)

Table 2. Teacher Characteristics

	Elementary	Secondary
Average Number of Years of Teaching Experience:	12.03 (7.50)	2.69 (7.50)
Average Number of Years of Teaching in District:	9.68 (6.94)	10.04 (7.00)
Percentage of Teachers whose Highest Diploma is:		
Bachelors	57.92	50.66
Masters	41.65	48.44
Doctorate	0.17	0.91
Average Number of credits/ units beyond last degree:	24.10 (24.39)	25.82 (22.34)
Average Number of students in class	27.11 (9.45)	26.09 (9.84)
Average Hours per week of Reading or Math:	6.55 (1.97)	5.38 (1.78)
Average Hours per week of Mathematics	5.19 (1.44)	5.62 (1.67)

CHAPTER 2

ASSESSING STUDENT ACHIEVEMENT: THE FREQUENCY OF TESTING AND THE TIME IT TAKES

As CSE researchers interviewed teachers across the United States, they spoke of the many ways in which they assess students' progress and monitor the results of their teaching. Routine class and homework assignments, teachers pointed out, provide recurrent information on students' learning. Classroom interaction — during question-and-answer recitation and discussions, when students ask for help with their work, as they read orally or work problems at the board, etc. — yields immediate, continuous feedback on how students are doing. Special projects, presentations, and reports offer additional data on student progress and teaching effectiveness. Testing, then, is viewed by teachers as only one among the many strategies in their repertoire for measuring students' achievement.

Teachers' interview remarks imply that *testing* means for them eliciting information from individual students, usually through paper-and-pencil instruments, under controlled conditions, i.e., conditions which preclude students' access to texts, notes, and others' assistance. While this definition of testing is hardly unique, it does differentiate teachers' view of testing from their perspective on assessment in general. From their viewpoint (as noted above), *assessment* of student achievement goes on constantly during the course of classroom teaching and learning. Testing, in contrast, occurs periodically in time set aside explicitly for that purpose. The amount of testing that teachers report thus represents only a small proportion of their assessment effects, an observation which provides important context for interpreting the following discussion on how much testing goes on in schools.

CSE's national survey asked teachers to list each type of test their students receive over the course of a school year in reading or English and mathematics, the frequency with which each type is administered to their "typical student," and the approximate length of time it takes that student to complete a usual test of each type. Teachers' responses provide a picture of the annual class time students spend taking tests in these basic skills subjects. This picture is described first in the sections below, then it is supplemented with fieldwork findings that highlight some additional time testing entails for both students and their teachers.

The National Picture: Modest Amounts of Time on Testing

Elementary students spend less than 10 percent of the annual allocated instructional time in basic skills testing. Table 3 shows the average annual time students devote to test taking, as well as the average frequency and duration of testing, in each subject and level of schooling surveyed.

As these figures indicate, the typical student in the upper-elementary grades spends about 10 hours a year taking reading tests and 12½ hours a year taking mathematics tests. Test taking, then, consumes about four percent of the average time allocated to formal instruction in reading and close to seven percent of the average time given to formal instruction in mathematics during the entire school year. (These percentages are based on the average instructional time reported by the elementary-school teachers surveyed: 6½ hours a week in reading, 5 hours a week in mathematics. Here and throughout this section, calculations assume a school year of 37 weeks or 180 days of actual instruction.)

Table 3. Time Devoted to Testing in Typical Classes

	Total Amount of Class Time Spent on Testing per Annum	No. of Test Sessions for Typical Student	Average Length of Session
Elementary School (Grades 4-6) — Reading Tests	9 hrs. 56 min.	22	27 min.
— Mathematics Tests	12 hrs. 28 min.	23	32 min.
10th Grade English Class	26 hrs. 34 min.	49	32 min.
10th Grade Mathematics Class	24 hrs. 18 min.	45	33 min.

Table 4. Time Devoted to Required Testing, As a Percentage of Total Testing Time For Typical Classes

	Percentage Time on Testing Required by State	Percentage Time on Testing Required by Local School District	Percentage Testing Time Devoted to Non-Required Tests
Elementary School (Grades 4-6)			
— Reading	30	29	41
— Mathematics	21	25	54
10th Grade English Class	12	13	74
10th Grade Mathematics Class	9	14	77

Elementary students take a test in reading and a test in math about once every eight days. Students' test-taking time, of course, is seldom distributed evenly from week to week across the school year. Periods or more intensive testing can occur at the elementary level, for example, during administration of placement and diagnostic measures, standardized test batteries (with their reading and math sub-tests), and end-of-book or end-of-level exams. Routine quizzes and chapter tests are often deferred at such times or in other special circumstances. With this caveat, the averages in Table 3 yield rough estimates of general testing patterns. They indicate that throughout the year the typical upper-elementary student faces a half-hour test in reading and a half-hour test in math about once in every eight school days.

High school students spend 12 to 13 percent of their time in English and mathematics class taking tests. Students in high school appear to spend more of their class time taking tests. Survey results reveal that the typical tenth-grader enrolled in an English class spends nearly 26½ hours yearly completing tests in that subject. This constitutes a little over 13% of their annual time in English instruction, which teachers' reports indicate averages 5.4 hours weekly across the school year.

A typical tenth-grade mathematics student devotes somewhat more than 24 hours to math tests in a school year. At an average of 5½ hours weekly for mathematics instruction, this equals about 12% of their class time.

High school students take an English test and a math test every three-to-four days. As Table 3 shows, in the subjects surveyed the average testing session in tenth grade last only moments longer than in upper-elementary classes. On the average, however, the typical tenth-grader is tested about twice as frequently. He or she encounters a half-hour test in English class roughly every three-and-a-half days; in mathematics class, about once every four days.

Mandated tests consume substantial proportions of students' total test-taking time. How much of the test-taking time just described results from tests mandated by agencies beyond the school? How much occurs at teachers' discretion? Table 4 provides answers to these questions. Elementary-school teachers in the sample report that on the average about half their students' test-taking time in both reading and math is spent on measures required by their state or school district. At the high-school level, state and district mandates account for about a quarter of the time students spend taking tests in both English and mathematics. Notice, then, that since high school students on the average spend twice as much time annually being tested as elementary students do, these percentages suggest that the actual number of hours spent in required testing is quite similar at both levels of schooling. Notice, too, that a greater proportion of assessment in the high school subjects is voluntary: conducted at the discretion of the individual teacher.

Students spend most of their time on teacher-developed tests. Which types of tests call for greater proportions of students' test-taking time? To address this question, the survey employed test-type categories that recurred consistently and spontaneously in the talk of teachers, school administrators, and counselors during open-ended pre-survey interviews. The goal was to give survey respondents a categorization system as similar as possible to the one they use naturally in their everyday thinking and conversation about assessment. As Table 5 demonstrates, this system differentiates tests primarily in terms of their point of origin, i.e., according to who develops the measure and/or requires its use.

Table 5. Time on Different Tests, As a Percentage of the Total Student Time Devoted to Test-Taking

TYPE OF TEST	Elementary Teachers		10th Grade English Teachers	10th Grade Mathematics Teachers
	Reading	Math		
Tests which form part of a statewide assessment program	3	3	5	1
Required Minimum Competency Tests	1	2	1	1
Tests included with curriculum materials	28	35	8	17
Other commercially published tests	17	18	6	3
Locally developed and district adopted tests	13	8	5	2
School or teacher developed tests	37	35	74	76

A glance at the results in Table 5 shows immediately that tests developed by individual teachers and schools and, at the elementary level, those which accompany commercial curriculum materials, occupy the great majority of students' testing time. Notice that these are the types over which teachers have most control. They can administer them when they deem appropriate; they can design (or readily adapt) the content to suit their own teaching emphases. Most teachers interviewed said that these types of tests fit best with their instructional schedules and curricula. And, from their points of view, these are the most valid instruments of those listed for such routine tasks as grading, on-going planning of teaching, etc. (This will be discussed further in Chapter 3). The predominance of locally developed tests at the secondary level supports the notion that high school teachers have more control over classroom assessment than do elementary school teachers. But heavy use of locally developed tests in the high schools may also reflect the limited number of suitable commercial testing materials available. Comprehensive curricular programs — including texts with coordinated workbooks, tests, etc. — are more widely available for teachers of the elementary grades.

Finally, note that the two types of testing most often generated by state policy — minimum competency testing and state assessment — consume on the average very small proportions of classroom testing time.

The figures in Table 5 are averaged across all teachers in the survey, including those in states without minimum competency testing requirements. Even where minimum competency tests (MCT) are required in the grades sampled, however, less than three percent of the testing time at the sampled elementary grade levels and two percent of the testing time in secondary grades and subjects sampled is taken up by these tests. Where MCT's are available, but not required, they absorb less than one percent of the total testing time in the grades and subjects surveyed.

The picture with regard to statewide assessment programs is similar. Such programs require no more than three percent of the total annual testing time at the elementary level (or about 45 minutes per year on the average for reading and mathematics combined). At the high school level, tenth grade English assessment programs typically take about 75 minutes annually and mathematics programs an average of 30 minutes per year.

Where there are no state minimum competency, proficiency, or functional literacy testing requirements, students spend more time on classroom achievement testing. Tests of minimum competency or proficiency or functional literacy are now required of all students in over 40 states, representing about two-thirds of the nation's student enrollment. In some states, passing these tests is a prerequisite for promotion to certain grades and/or for high-school graduation. In others, they are mandated only for diagnostic purposes: to assure that students with deficiencies in basic skills are identified and offered remedial instruction. Furthermore, some states designate specific instruments that must be used in minimum competency testing, while legislation in other states permits local school districts to select or construct tests of their own choice.

Teachers' reports suggest that these minimum competency requirements may somehow be affecting the amount of classroom achievement testing teachers otherwise do. At least, teachers' survey reports show that, when other sampling factors are controlled,* students in states

Table 6. Relationships Between State Minimum Competency Testing Requirements and Students' Test-Taking Time

Reported in Minutes

STATE REQUIREMENT	SECONDARY			ELEMENTARY		
	English	Math	Total per Teacher[1]	English	Math	Total per Teacher[2]
No Minimum Competency Testing (MCT)	3723.53	3173.38	3455.01	577.45	570.91	1148.37
MCT required for diagnosis, state mandated measure	915.77	1180.50	1086.47	504.32	488.15	922.48
MCT required for diagnosis, local choice of measure	1600.07	1394.57	1482.77	489.90	486.32	976.22
MCT required for promotion or graduation, state measure	1427.73	805.15	1095.86	388.69	632.88	971.57
MCT required for promotion or graduation, local choice of measure						

[1] Difference in mean values of different MCT categories statistically significant at $p < .01$.
[2] Difference in mean values not significant statistically.

---

*Other factors considered in sampling include districtwide socioeconomic status, district enrollment size, geographic region in the nation, and urban-suburban-rural locale. See the introduction for further details.

with no minimum competency requirements at all spend more time on achievement testing each year than students elsewhere do. (See Table 6.) This difference is dramatic (and statistically significant) at the secondary level, where all types of minimum competency requirements appear to be accompanied by much less classroom testing (from 33 to 45 hours less annually) and where competency requirements for promotion or graduation are accompanied by the least testing time of all.

At present, this pattern is difficult to explain. On the surface, it seems to suggest that teachers have eschewed routine classroom testing in favor of minimum competency measures: that they are permitting minimum competency tests to take place of other forms of assessment. This interpretation, however, makes little logical sense. Proficiency or minimum competency tests are given only at certain grade levels. Typically, too, they are given in those grades only on a single occasion. Thus, they cannot possibly supply the feedback on student performance that teachers need regularly for monitoring students' learning progress, assigning report card grades, making on-going teaching plans, and so on. Furthermore, fieldwork visits to various states with different minimum competency requirements revealed no reduction in routine tests and quizzes. In fact, fieldwork suggested that at least in the districts visited, additional time can be spent in testing to assure that students perform well on minimum competency measures. Nevertheless, careful review of the survey instruments and the statistical analyses to which they were subjected substantiates the findings displayed in Table 6. The processes that underlie and explain these results await further study.

Socioeconomic status (SES) seems unrelated to students' test-taking time. Given the evaluation and testing requirements that are commonly associated with compensatory education programs, and given that these programs serve students from lower socioeconomic backgrounds, many people have speculated that lower SES students spend more of their school time on testing than students from higher SES homes. CSE survey results, however, indicate that this is not the case. Students in lower SES areas do not spend more time taking tests than those in middle-income or upper-income settings, nor do they even spend more time taking tests required by their district, their state, or in conjunction with federal educational program guidelines. This finding holds true regardless of whether a district-level or a school-level indicator of socioeconomic status is used.

In concluding this section, it is also worth noting that no other variable included in this study (except minimum competency requirements) appeared to have any relationship with the amount of time students spend taking tests.

The discussion so far has centered on how much testing goes on in the basic-skills subjects of reading or English and mathematics across the nation's schools. Emphasis has been on the frequency of testing and on the class time students spend with tests in hand, actually completing them. Survey questions purposely focused on these topics as especially relevant to a portrait of national practices.* Fieldwork results elaborate these findings, providing an illustrative look at all the time students spend on testing, at teachers' testing time, and at time on testing across the curriculum.

Testing consumes student time before and after the test. In most classrooms, testing demands more class time than that required for students to complete their tests — time which is spent both before and after they answer test questions. Wide-ranging interviews with teachers, conducted by CSE both before and after the national survey, illustrate how this time is spent and how much it can add up to.

Preparations for testing can begin days or even weeks before the test is given. At a minimum, teachers inform their students when the test will be, explain what it will cover, and say a word or two about the question formats that students can expect. When mandated measures such as standardized batteries or minimum-competency tests are due, however, some teachers spend class time to train students in their specific response formats and/or in general test-taking strategies. Some also suspend teaching of the on-going curriculum, devoting class time instead to review and practice of skills and content that they know these tests will cover.

When the testing day arrives, of course, time is required for passing out materials, giving directions, and handling students' questions. In order to provide an appropriate environment for testing, some teachers say, they routinely allow several moments for "settling students down" and/or rearranging students' seating. Filling in student-identification information and covering directions can be especially time-consuming at the outset of special testing episodes. At the elementary level, teachers often report spending a half-hour or more on these preliminaries when standardized testing, state assessment, or minimum competency measures are administered. Moving students from their classrooms to special testing locations (the library, cafeteria, etc), as is sometimes done for the latter types of assessment and for high-school finals, is another before-testing activity that can take up time.

---

*In addition, project resources were insufficient to examine testing in all subject areas, and both pre-survey interviews and questionnaire piloting confirmed that eliciting information on all the time associated with preparing for, taking, and reviewing tests would place an enormous response burden on survey recipients.

Once students have completed a test, class time is given over to collecting papers. Sometimes, tests are corrected in class. Then, if necessary, regular classroom seating patterns are restored. Nearly all teachers in the elementary grades report that they regularly set aside time for students to "relax" or "cool out" after particularly important or lengthy examinations. Some high schools accomplish this with special schoolwide schedules for finals and (less often) mid-terms.

The amount of class time such activities as these consume appears to vary markedly from classroom to classroom and school to school. In two elementary schools, for example, every teacher in grades K through 6 was interviewed about all the time their students spend on test-related activities in all subjects throughout the school year. In one of these schools (Hillview Elementary), students usually spend an average of 91% of their total, testing-related time actually answering test questions. Only 9%, on the average, of the typical student's total time on testing each year is taken up with before-the-test and after-the-test activities of the kind described above. In the second elementary school (Cityside), however, much more time is routinely spent on pre-testing drills and review which, teachers avowed, were undertaken only because mandated testing was about to occur. Furthermore, logistics in support of testing — scheduling changes that reduced class time; room reassignment for testing, etc. — claims a great deal of instructional time during required-test administration each spring in this densely populated school. Thus, students here spend only 55% of the average annual time devoted to test related activites actually taking tests. They devote nearly as much time each year, in other words, to before-the-test and after-the-test activities as they do to test taking. (For details on these two schools, their testing programs, and their districts' testing programs, etc., see Dorr-Bremme, 1983.)

Similar interviews were conducted, although less intensively in any one school, with high school teachers. These suggest that secondary students usually spend 10 to 15 percent of their total yearly testing time in any one class on before- and after-testing activities.

The percentages offered here, of course, are only illustrative. Nevertheless, they do provide a useful context for interpreting the national averages of students' test-taking time cited earlier.

In two elementary schools, testing across the curriculum consumed eight to ten percent of students' available instructional time. How much time do students spend on all test-and-testing related activities in subjects across the curriculum? Fieldwork interviews in the two schools mentioned in the last section also provide illustrative answers to this question for students in elementary school. In the first of two schools

(Hillview), for instance, an average student devotes 88 hours a year preparing for, taking, and winding up and going over tests in all subjects. This comprises about 10% of their annual class time (which equals five hours daily, excluding lunchtime and recess, over 177 school days, or 885 hours per year). Across classrooms in the other elementary school cited above (Cityside), students' total testing time in all subjects averages 76 hours a year, or 8.6% of their annual class time of 885 hours. Observations of testing episodes — including the before, during, and after phases — suggest that the interview estimates upon which these totals are based are generally quite accurate.

Tables 7 and 8 show how this time is distributed by subject area. Notice that all teachers do not test in all subjects and that testing in the basic skills subjects of reading and mathematics (not including multi-subject batteries which also cover these subjects) consumes about 50% of students' total time on testing in these two schools.

For each hour that students spend taking tests, teachers seem to spend two-to-three more. The annual times students spend on test-taking (Table 3 above) can serve as a rough indicator of the times that teachers spend giving tests in the classroom. CSE's interviews with teachers confirm that in most cases teachers actively monitor the class and answer students' questions as testing is in progress. These same interviews, however, suggest that teachers spend only about a quarter to a third of their total time on testing in this way. That is, for each hour they devote to giving a reading or math test, they typically spend another two or three hours on such activities as preparing for testing (e.g., constructing tests and dittoing them, reviewing directions for state assessment or standardized-test administration), correcting and grading tests, recording scores, etc. At the elementary level, teachers also find that they spend a good deal of time checking over special answer sheets used for machine scoring to be sure that the identification information is correct, that there are no stray pencil marks to throw off the scoring, etc.

Interviews with elementary-school teachers indicate that they spend about 12 to 15 percent of their annual reported work time, both in and out of school, on achievement testing in all subject areas. This averages about 200 to 250 hours through a school year. (Similar figures are unavailable for high-school teachers, but they do appear to spend two hours or so outside of class for every class hour of student testing.)

Tables 7 and 8 also display the total time on testing that teachers in the two case study elementary schools (Hillview and Cityside) spend annually on testing in each subject. Note that testing in reading and mathematics together demands over 50 percent of the total teacher time on testing at each school. If the testing in these subjects that takes place as part of multi-subject batteries were included, this percentage would be higher.

Other staff members' time on testing. Administrators, as well as classroom aides (or paraprofessionals) and volunteers, also play a role in the work of testing. Classroom assistants spend their time much as teachers do: proctoring test administration, grading tests and recording scores, etc. School administrators typically spend their time coordinating major schoolwide testing programs: overseeing distribution, administration, collection and checking of state-assessment measures, standardized testing, and/or minimum-competency (proficiency) assessment. (See Tables 7 and 8 for the time administrators and classroom assistants spend annually on all aspects of testing in the two case study schools.)

Table 7. Hillview School — Littleton District
Distribution of Staff & Student Testing Time

By Subject

Each staff category cell shows:
- No. of staff members involved
- Avg. hours/staff member/year
- % Total testing time for staff category

SUBJECT AREAS	ADMINISTRATORS' TIME	CLASSROOM TEACHERS' TIME	INSTRUCTIONAL SPECIALISTS' TIME	VOLUNTEERS' TIME	TOTAL STAFF TIME (In Person Hours)	AVG. STUDENT TIME PER STUDENT (hours)	NUMBER OF CLASSROOMS Total = 30
Reading		11 52.47 20.7%	1 17.4 8.8%	1 5.0 6.4%	599.6 19.0%	12.12	11
Mathematics		11 77.11 30.5%	1 53.9 27.3%	3 15.44 59.7%	948.46 30.0%	25.11	11
Language Arts		8 24.30 7.0%	1 34.75 17.6%		229.17 7.3%	7.81	8
Spelling		8 51.42 14.8%	1 21.58 10.9%		432.97 13.7%	19.34	8
Social Studies		5 19.55 3.5%			97.75 3.1%	4.53	5
Science		5 28.0 5.0%			140.0 4.4%	5.8	5
Health — Phys. Ed		3 8.33 0.9%			25.0 0.8%	7.19	3
Other, Miscellaneous		3 8.61 1.0%	1 70.0 35.4%		95.83 3.0%	3.39	3
Multi-Subject*	2 49.87 100.0%	11 42.06 16.6%		3 8.78 33.9%	5[illegible]8.77 18.6%	23.93	11
TOTALS By Staff category (In person hours)	99.75 100.0%	2782.5 100.0%	197.63 100.0%	77.66 100.0%	3157.55 99.9%		

*The Multi-subject category includes standardized tets which assess performance in several subject areas. Also included in this category is the general intelligence test given twice a year at the same time as (i.e., on a day contiguous with) the standardized test. Some respondents reported time devoted to the intelligence tests as separate from that given to the standardized test; others did not. Thus, time devoted to both is collapsed here.

Table 8. Cityside School — Metro District
Distribution of Staff & Student Testing Time

By Subject

Each staff category cell shows:
- No. of staff members involved
- Avg. hours/staff member/year
- % Total testing time for staff category

SUBJECT AREAS	ADMINISTRATORS' TIME	CLERICAL TIME	CLASSROOM TEACHERS' TIME	INSTRUCTIONAL SPECIALISTS' TIME	AIDES' (Para-professionals) TIME	VOLUNTEERS' TIME	TOTAL STAFF TIME (In Person Hours)	AVG. STUDENT TIME PER STUDENT (hours)	NUMBER OF CLASSROOMS Total = 30
Reading	2 139.66 74.5%	1 10.3 100.0%	28 54.61 25.6%	1 74.0 45.0%	26 15.31 30.7%	1 11.67 12.6%	2302.42 28.8%	9.43	28
Mathematics			27 67.58 30.5%		25 15.51 29.9%	2 33.06 71.8%	2278.38 28.6%	21.01	27
Language Arts			16 25.42 6.8%		10 3.63 2.8%		443.0 5.5%	18.71	16
Spelling			22 54.25 20.0%		18 11.17 15.5%	1 9.17 10.0%	1403.67 17.6%	25.83	22
Social Studies			10 17.65 2.9%		6 4.12 1.9%		201.20 2.6%	10.33	10
Science			5 16.4 1.4%		2 0.63 0.09%		83.25 1.0%	4.33	5
Health — Phys. Ed			6 16.55 1.7%		6 9.52 4.4%		156.47 2.0%	30.28	6
Other, Miscellaneous			6 40.27 4.0%	1 74.0 45.0%	4 10.34 3.2%		356.96 4.5%	0.39	6
Multi-Subject	3 31.90 25.5%		26 16.24 7.1%	2 8.16 10.0%	28 5.39 11.6%	2 2.6 5.6%	690.45 9.4%	9.62	26
TOTALS By staff category (In person hours)	375.0 100.0%	10.3 100.0%	5975.32 100.0%	164.33 100.0%	1298.5 100.09%	92.22 100.0%	7915.8		

CHAPTER 3

USING ASSESSMENT RESULTS

The results of tests and other assessment techniques can be used for many different purposes by educators in the schools. Nearly all educational testing and measurement texts include long lists of these: diagnosing learners' needs, placing students in programs, monitoring students' progress, evaluating curriculum and instruction, planning for school improvement, reporting to parents, satisifying accountability requirements, and many others. Such lists outline the possibilities. CSE's Test Use in Schools Study sought to identify actual practices. Thus, both principals and teachers were asked how heavily they weigh different types of test results and information from other sources in a variety of routine decisions and tasks.

Figure 2, an example from the teacher survey, illustrates the form these questions took.

Figure 2

Format of Survey Test-Use Questions for Teachers and Principals
Illustration from the Teacher Survey

22. When I initially group or place students for instruction, here's how important various sources of information are to me:

	Crucial	Important	Slightly important	Unimportant	Not available
(a) Previous teacher's comments, reports, grades	4	3	2	1	0
(b) Students' standardized test scores	4	3	2	1	0
(c) Students' scores on district continuum or minimum competency tests	4	3	2	1	0
(d) Results of placement tests included with curriculum use	4	3	2	1	0
(e) Results of other special placement tests	4	3	2	1	0
(f) Results of tests I make up	4	3	2	1	0
(g) My own observations and students' classwork	4	3	2	1	0

The same format was followed in the questionnaires for principals. As in the example, each question about a particular use of assessment elicited information about a range of test types and about other modes of assessment, e.g., observations and classwork, as well. Notice that the test-type categories given in these questions are identical with those employed in survey questions about students' testing time (Table 5 above). Recall that these were the test-type labels teachers and principals used recurrently, without prompting, during the open-ended, pre-survey interviews conducted in several school districts across the United States. It is highly likely, therefore, that most survey respondents found them familiar and meaningful.

Practically, the survey could not examine all the possible school and classroom uses of assessment results. Choices had to be made in order to keep questionnaires at a reasonable length. Pre-survey interviews played a major role in guiding these choices. One of these interviews asked respondents to name all the achievement tests that they gave their students through the school year, then to describe what (if anything) they did with the results. The second interview form encouraged informants to discuss the major tasks and decisions their jobs routinely entailed as a typical school year proceeded; it then inquired about all the information that informed each task and decision. These interviews made it possible to identify: (1) those tasks and decisions that teachers and principals considered to be major responsibilities in their respective jobs; and (2) those for which principals or teachers were inclined to consult test scores or other assessment information. Thus, within space contraints, the survey questionnaires were able to focus on major tasks and decisions in which test results were likely to be used.

Below, the findings from the principals and teachers questionnaires are described and discussed separately, then supplemented with information from fieldwork interviews.

A Wide Variety of Assessment Results Play a Role In School-Level Tasks, But Teachers' Tests and Their Professional Judgments Are Most Important.

Principals described the importance of different types of assessment results in eight, school-level tasks and decisions. Table 9 lists these and shows the percentages of principals who stated that the different types of assessment information were crucial or important in each task. Table 10 displays the same data in a different form: as the mean (or average) importance rating principals gave each type of information for each task.

Notice that both tables report the use of five main types of assessment results: those that come from (1) standardized, norm-referenced batteries; (2) minimum competency (proficiency) tests; (3) tests referenced to district curriculum objectives; (4) teachers' classroom tests and assignments (unit or chapter tests, quizzes, finals, whether teacher-constructed or included with published curriculum materials); and (5) teachers' observations of and interactions with students and/or their professional judgments. In fact, however, principals were also asked to rate the importance of other types of information for five of the eight tasks.

Table 9. School-Level Uses of Test Results and Other Information (Percentages of *Principals* Reporting Use of This Information as Crucial or Important for the Specified purpose)

Task or Decision	Information Source					
	A	B	C	D	E	F
			ELEMENTARY			
Curriculum Planning	78	60	65	72	88	—
Assigning Students to Classes	47	30	38	74	84	49[a]
Teacher Evaluation	16	11	25	40	—	100[b]
Allocating Funds	28	21	29	—	81	77[c]
Student Promotion	51	36	48	84	96	94[d]
Informing the Public	72	38	41	42	—	—
Communicating to Parents	78	56	63	98	95	92[e]
Reporting to District	81	55	58	53	—	—
			SECONDARY			
Curriculum Planning	74	75	57	63	84	—
Assigning Students to Classes	72	64	45	75	80	76[f]
Teacher Evaluation	20	15	21	43	—	95[b]
Allocating Funds	24	28	21	—	94	84[c]
Student Promotion	24	48	26	84	76	96[f]
Informing the Public	74	63	43	47	—	—
Communicating to Parents	79	69	45	96	94	97[f]
Reporting to District	86	72	56	60	—	—

A = Results of standardized, norm-referenced batteries
B = Results of minimum competency (proficiency) tests
C = Results of district's, objectives-based tests
D = Results of teachers' classroom tests and assignments
E = Teacher' opinions, judgments, recommendations
F = Various other sources, as follows:
- a = students' past classroom behavior
- b = observations of teachers' teaching
- c = specific directions from district
- d = classwork throughout the year
- e = observations of the student
- f = student's report card grades

— = Not asked

Table 10. Importance of Test Results and Other Information In School-Level Tasks and Decisions (Mean Ratings by *Principals* on a Four-Point Scale)*

Decision or Task	A	B	C	D**	E	F
			ELEMENTARY			
Curriculum Planning	3.01 +	2.91	3.04	2.99	2.94	3.27
	(.67)	(.75)	(.87)	(.07)	(.84)	(.64)
Assigning Students to Classes	2.50	2.35	2.46	2.44	2.93	3.12
	(.81)	(.91)	(.99)	(.08)	(.79)	(.71)
Teacher Evaluation	1.70	1.53	1.80	1.68	2.12	—
	(.76)	(.78)	(.93)	(.14)	(.97)	
Allocating Funds	1.91	1.89	1.94	1.91	—	3.08
	(.87)	(.90)	(1.01)	(.03)		(.71)
Student Promotion	2.65	2.31	2.38	2.45	3.05	3.29
	(.81)	(.96)	(.94)	(.18)	(.70)	(.67)
Informing the Public	2.77	2.47	2.34	2.52	2.31	—
	(.90)	(.99)	(1.00)	(.22)	(1.05)	
Communicating to Parents	2.91	2.64	2.67	2.74	3.43	3.45
	(.60)	(.98)	(.95)	(.15)	(.55)	(.57)
Reporting to District	3.12	2.78	2.74	2.88	2.62	—
	(.68)	(1.10)	(1.10)	(.21)	(.91)	
			SECONDARY			
Curriculum Planning	2.83	3.27	2.95	3.02	2.76	3.14
	(.67)	(.64)	(.82)	(.23)	(.75)	(.70)
Assigning Students to Classes	2.77	2.98*	2.78	2.84	2.98	2.99
	(.77)	(.87)	(.87)	(.12)	(.73)	(.79)
Teacher Evaluation	1.63	1.77	1.84	1.75	2.39	—
	(.74)	(.71)	(.78)	(.11)	(.83)	
Allocating Funds	1.73	2.20	2.06	2.00	—	3.34
	(.81)	(1.13)	(1.08)	(.24)		(.54)
Student Promotion	1.61	2.58	2.05	2.08	3.33	3.46
	(.78)	(1.28)	(1.13)	(.49)	(.85)	(.75)
Informing the Public	2.84	2.92	2.30	2.69	2.24	—
	(.80)	(1.03)	(1.07)	(.34)	(1.05)	
Communicating to Parents	2.91	3.03	2.55	2.83	3.56	3.38
	(.58)	(1.00)	(.99)	(.25)	(.55)	(.76)
Reporting to District	3.10	3.12	2.92	3.04	2.53	—
	(.64)	(.97)	(.95)	(.11)	(.88)	

A = Standardized, norm-referenced test batteries
B = Minimum Competency Tests
C = District Objective-based Tests
D = Average Required Tests (A,B,C)
E = Results of Teacher and Curriculum tests
F = Teacher Opinions/Recommendations
* [4-point scale: 4 = Crucial Importance — 1 — Unimportant or not used]
+ Numbers in parentheses are standard deviations.
** Numbers in parentheses are standard deviations of values in columns A, B and C.

Table 9 (Column F) shows which of these other types of information most principals considered crucial or important for each of those five tasks, as well as the percentages who did so. For the sake of simplicity, these data are omitted in Table 10.

As the tables indicate, most schools appear to ground their actions upon several information sources in all eight tasks or decisions. In general (Table 10), no one stands out as markedly more important than all the others for most tasks. For almost every task, however, principals rate the results of teachers' classroom testing as more crucial or important more often than the results of any other type of paper-and-pencil measure (See Table 9). What is more, for each of the eight tasks listed, teachers' opinions, judgments, and recommendations clearly carry more weight than any type of test results.

Some types of measures listed on the survey are more formal tests: standardized, norm-referenced batteries, other kinds of minimum competency measures,* and test referenced to districts' instructional objectives. Compared to teacher-made tests and class assignments, great attention is usually given to their psychometric quality and their administration is usually marked by more formal or "official" testing arrangements and procedures. Usually, too, these tests are given in schools at the mandate of an agency beyond the school, e.g., by the district, the state or, even by the federal government as part of the requirements for a specially funded program.

The results of these formal tests appear to make their greatest contribution in three school-level tasks: curriculum planning, communicating to parents about their children's achievement, and reporting to school district administrators. Conversely, formal test results are least important in evaluating teachers and in allocating funds within the school for such things as personnnel, equipment, and materials. In secondary schools, formal test results, and especially the results of minimum competency or proficiency tests, also play a significant role in decisions about students' class assignments. Fieldwork indicates, for example, that students who fail to meet minimum standards on competency tests are sometimes assigned to special courses designed for remediation in the basic skills covered by the tests.

Standardized, norm-referenced batteries seem to be the most influential of the formal required tests at the elementary level. However at the high school level, educators pay more attention to the results of minimum competency tests than to those of the other types of formal measures.

---

*In some states and districts, standardized, norm-referenced measures are used as minimum competency or proficiency tests.

The Results of Formal Tests Are Deemed More Important In Schools Serving Students of Lower Socioeconomic Status (SESS)

An earlier section (page 21) noted that students in lower SES schools do *not* spend more time taking tests than middle or upper-income pupils do. Furthermore, teachers' classroom uses of test results (to be discussed next) do not vary systematically or significantly with students' socioeconomic status. In schoolwide or school-level tasks and decisions, however, tests results *do* appear to have greater impact and wider consequences in lower SES schools than they do in higher SES settings. In the former, principals report that more importance is accorded the scores of formal tests — especially minimum competency measures and district objectives-based tests — in planning curriculum, deciding on students' class assignments, allocating school funds, and reporting on school achievement to the public-at-large, parents, and district officials. (See Table 11, which shows the results for all principals, elementary and secondary together, divided into higher and lower SES groups using school-level indicators.)

For Classroom Tasks, Teachers Place Most Weight on Their Observations and the Results of Their Own Tests.

Teachers were asked to rate the importance of the results of various assessment types in four routine classroom tasks or decisions. The proportions of elementary and high school teachers who described different types of results as crucial or important in each is displayed in Tables 12 and 13. Table 14 portrays similar data in a different form: as the mean (or average) rating teachers gave each type of information for each of the four tasks. Notice that Tables 12 and 13 divide teachers' responses by subject matter, while Table 14 does not.

These tables demonstrate that teachers *do* use test results of various types in making common instructional decisions. They also reveal quite clearly, however, that teachers place greatest trust in their own observations of students' class performance and in their personal, clinical judgment. Nearly every teacher reporting says that their "own observations and students' classroom work" are crucial or important sources of information for initially grouping or placing students, in deciding to change students' placement or grouping, and in determining students' report-card grades. The great majority also give heavy weight to the results of their own, self-constructed test in each of these tasks. Among teachers in the elementary grades, "the results of tests included with the

Table 11. Importance of Test Results for School Decisionmaking in Schools of Higher and Lower Socioeconomic Status (SES)*

Decision Or Task	Standardized Norm-referenced Test Batteries	Minimum Competency Tests	District Objectives based or Continuum Tests	Average Required Tests (A,B,C)
		HIGHER SES		
Curriculum Evaluation	2.90 (.52)	2.95 (.71)	2.64 (.92)	2.83
Student Class Assignments	2.49 (.71)	2.24 (.79)	2.10 (.96)	2.27
Teacher Evaluation	1.69 (.72)	1.81 (.74)	1.94 (.81)	1.81
Allocating Funds	1.85 (.83)	1.85 (.91)	1.71 (.86)	1.80
Student Promotion	2.19 (.83)	2.49 (1.04)	2.27 (.95)	2.31
Public Communication	2.69 (.78)	2.36 (.96)	2.33 (1.00)	2.46
Communicating to Parents	2.80 (.56)	2.74 (.94)	2.51 (.84)	2.68
Reporting to District	3.03 (.73)	2.94 (1.09)	2.74 (.94)	2.90
		LOWER SES		
Curriculum Evaluation	3.08 (.78)	3.18 (.59)	3.08 (.83)	3.11
Student Class Assignments	2.68 (.79)	2.67 (1.03)	2.59 (.94)	2.65
Teacher Evaluation	1.95 (.84)	1.74 (.72)	1.94 (1.03)	1.88
Allocating Funds	2.00 (.79)	2.45 (.92)	2.18 (1.00)	2.21*
Student Promotion	2.45 (.93)	2.39 (.99)	2.17 (.84)	2.34
Public Communication	2.84 (.90)	2.93 (.97)	2.59 (1.04)	2.79
Communicating to Parents	2.96 (.57)	3.26 (.78)	3.26 (.51)	3.16
Reporting to District	3.11 (.65)	3.28 (.61)	3.11 (.93)	3.17

*[4-point scale: 4 = Crucial Importance — 1 = Unimportant or not used]

Table 12. Classroom Uses of Test Results and Other Information: (Percentages of ELEMENTARY teachers surveyed reporting use of this information as crucial or important for the specified purpose)

	Planning Teaching at Beginning of School Year		Initial Grouping or Placement of Students		Changing a Student from One Group or Curriculum to Another		Deciding on Students' Report Card Grades	
Source/Kind of Information	**Reading**	**Math**	**Reading**	**Math**	**Reading**	**Math**	**Reading**	**Math**
Previous teachers' comments, reports, grades	57	52	62	55	x	x	x	x
Students' standardized test scores	57	54	57	52	55	53	17	16
Students' scores on district continuum or minimum competency tests	51	47	50	45	45	39	20	18
My previous teaching experience	94	94	x	x	x	x	x	x
Results of tests included with curriculum being used	x	x	78	67	83	82	75	77
Results of other special placement tests	x	x	61	56	x	x	x	x
Results of special tests developed or chosen by my school	x	x	x	x	56	52	42	42
Results of tests I make up	x	x	80	86	78	85	92	95
My own observations and students' classroom work	x	x	96	97	99	99	98	98

Table 13. Classroom Uses of Test Results and Other Information: (Percentages of SECONDARY teachers surveyed reporting use of this information as crucial or important for the specified purpose)

	Planning Teaching at Beginning of School Year		Initial Grouping or Placement of Students		Changing a Student from One Group or Curriculum to Another		Deciding on Students' Report Card Grades	
Source/Kind of Information	**English**	**Math**	**English**	**Math**	**English**	**Math**	**English**	**Math**
Previous teachers' comments, reports, grades	28	29	34	40	x	x	x	x
Students' standardized test scores	47	29	49	30	62	39	12	8
Students' scores on district continuum or minimum competency tests	48	30	47	36	53	36	9	5
My previous teaching experience	99	97	x	x	x	x	x	x
Results of tests included with curriculum being used	x	x	45	35	58	43	44	31
Results of other special placement tests	x	x	42	26	x	x	x	x
Results of special tests developed or chosen by my school	x	x	x	x	50	31	28	34
Results of tests I make up	x	x	87	77	92	91	99	99
My own observations and students' classroom work	x	x	99	93	99	97	99	95

curriculum being used" play a major role in these same tasks. Notice, too, that teachers at both levels of schooling count their own, previous experience as teachers as their most important source of information for planning teaching at the beginning of a school year or semester.

Mirroring findings for principals, these results show that teachers place less emphasis on formal test results than they do upon information they gather themselves. Nevertheless, teachers do rate formal test scores as somewhat important (Table 14) for initial planning and placement decisions, as well as in deciding later on to reassign individual pupils to a different group or curriculum. Fieldwork indicates that in the latter process, teachers frequently treat test results as a general indicator of the students' "capabilities." Teachers interviewed said that they might examine standardized test scores, for example, to see if a poorly performing student has "low ability" or "isn't working up to his ability level." High-school interviewees sometimes explained that they checked the test scores printed on their class enrollment lists (as one put it) "to be sure they really belong in this class."

The data in Tables 12, 13, and 14 hint that teachers rarely rely on only one type of assessment information as they go about making instructional decisions. Table 15 confirms that for many this is in fact the case. Not only do a good number of teachers routinely consult several types of assessment results in reaching each decision listed, they consider many as equally crucial or important. This tendency is especially common among elementary teachers in the sample.

Table 16 elaborates on this last point and, in effect, summarizes the key points of the discussion in this section. It demonstrates that except in planning their teaching at the beginning of a school year or semester, only small proportions of teachers count one source of assessment information as more important than all others for the routine tasks listed. And of those teachers who do report trusting one kind of information above all the rest, from 86 to 100 percent say that the information they trust most is their own observations and students' classwork (or, in the case of planning at the start of the year, their previous teaching experience).

Table 14. Importance of Test Results and Other Information In Classroom Tasks and Decisions

(Mean Ratings by Teachers on a Four-Point Scale)*

Decision Area:	Standardized Test Batteries	District Continuum or Minimum Competency Tests	Tests Included with Curriculum	Teacher-Made Tests	Teacher Observations/ Opinions
			ELEMENTARY		
Planning teaching at beginning of the school year	2.53 (0.74)	2.60 (0.79)	—	—	3.39 (0.76)
Initial grouping or placement of students	2.51 (0.74)	2.59 (0.82)	2.91 (0.74)	3.12 (0.83)	3.58 (0.78)
Changing a student from one group or curriculum to another, providing remedial or accelerated work	2.52 (0.79)	2.52 (0.81)	3.04 (0.74)	3.12 (0.84)	3.66 (0.72)
Deciding on report card grades	1.62 (0.76)	1.81 (0.81)	2.89 (0.79)	3.38 (0.74)	3.69 (0.72)
			SECONDARY		
Planning teaching at the beginning of the school year	2.22 (0.84)	2.38 (0.93)	—	—	3.59 (0.60)
Initial grouping or placement of students	2.28 (0.92)	2.46 (0.98)	2.48 (0.92)	3.04 (0.87)	3.84 (0.85)
Changing students from one group or curriculum to another, providing remedial or accelerated work	2.52 (0.95)	2.59 (0.86)	2.67 (0.93)	3.27 (0.76)	3.61 (0.66)
Deciding on report card grades	1.36 (0.66)	1.45 (0.64)	2.29 (0.96)	3.65 (0.62)	3.68 (0.65)

* [4-point scale: 4 = Crucial Importance — 1 = Unimportant or not used]

Table 15. Proportion of Teachers Who Report Considering Many Types of Assessment Information Critical/Important for Given Tasks

	Planning Teaching at Beginning of School Year	Initial Grouping or Placement of Students	Changing Grouping or Placement	Deciding on Report Card Grades
Number of Sources of Information Given in Question on Survey	4	7	6	6
Number of Sources Defined as "Many" for Purposes of this Analysis	3	4	4	4
Proportion of *Elementary Teachers* Who Indicated That at Least This Many Functioned as Critical and/or Important for the Given Activity	50%	71%	62%	40%
Proportion of *High School* Teachers	33%	47%	49%	20%

Table 16. Percentages of Teachers Who Consider One Type of Assessment Information To Be More Important Than Any Other

	ELEMENTARY		SECONDARY	
Task or Decision	**% of Total**	**% choosing teacher observation/judgment* as most important**	**% of Total**	**% choosing teacher observation/judgment* as most important**
Planning teaching at the beginning of the school year	48	89	68	97
Initial grouping or placement of studies	25	88	36	92
Changing a student from one group or curriculum to another	27	88	25	86
Deciding on students' report card grades	21	91	10	100

*Percentages in these columns are the percentages of those teachers who *did select one type of assessment as more important than all the others*, rather than percentages of all teachers in sample.

Fieldwork Interviews Support and Elaborate Survey Findings.

In the on-site interviews, teachers were able to describe with minimal constraints how they used test results and information from other assessment techniques. The purposes they most frequently cited were those that constitute their most essential, routine work: deciding what to teach and how to teach it to students of different achievement levels; keeping track of how students are progressing and how they (the teachers) can appropriately adjust their teaching; and evaluating and grading students on their performance (see Table 17). Clearly, these are the day-to-day routines of teaching.

Less frequently, respondents mentioned using assessment results in deciding to refer students who need special instruction and to counsel, advise, and direct students. These are important teaching responsibilities, but ones that serve to support or facilitate more basic instructional work.

Interviews also show that, unconstrained by the response format of the questionnaire, teachers still indicate that all types of paper-and-pencil measures they have available for assessing students' achievement, they rely most often on those that they themselves develop. As Table 17 shows, teachers freely cited more uses for such assessment tools than for any of the other types. The teachers' interviewed universally reported that their own perceptions' of children's performance in class, or homework, etc. were an important factor in all their judgments and decisions; thus the frequency with which these were mentioned is not included in Table 17. Fieldwork findings, then, are completely consonant with survey results despite differences in the elicitation procedures.

Fieldwork interviews also help to explain some of the reasons why teachers feel that the results of one type of test, or even of tests in general, cannot be trusted without reference to their everyday experience with learners. The following quotations are illustrative:

- I don't rely heavily on a lot of the test scores because I find that . . . some students are test takers and others are not . . . some students can handle the format, the time limit, (but in many cases) students are capable of more than the test scores show.
- I hate to say it, but I'd say about a third of these students don't give it their best shot. They feel there's nothing in it for them. There's no grade for it; there's no use for it — so they don't care.
- If I see there are certain kids having trouble I may look at their folders and find out (more) about them. But I try not to be swayed by somebody else's judgment . . . I may get more out of them by

Table 17. Types of Tests and the Uses of Their Results by Teachers (Interview Data)

(Cells show the number of times the 44 interviewed teachers freely cited each use for each type of test)

USES	TEST TYPES									
	A	B	C	D	E	F	G	H	I	Total
Planning Instruction	24	21	10	3	2	3	13	4	2	82
Referral/Placement	3	6	0	2	0	0	11	1	0	23
Within Classroom Grouping & Individual Placement	6	14	18	4	6	5	4	3	1	61
Holding Students Accountable for Work, Discipline	8	2	3	0	0	0	0	0	0	13
Assigning Grades	32	8	17	5	1	1	1	1	0	66
Monitoring Students' Progress	18	12	17	2	1	1	0	1	0	51
Counseling & Guiding Students	10	6	0	0	1	2	3	0	0	22
Informing Parents	0	1	0	0	0	1	0	0	0	2
Reporting to District Officials, School Board, etc	0	3	1	0	0	2	0	0	0	6
Comparing Groups of Students, Schools, etc	0	1	0	0	0	1	1	0	0	3
Certifying Minimum Competency	0	0	0	0	0	0	0	1	0	1
TOTAL USE CITATIONS	101	74	63	16	11	19	33	10	3	330
Explicit Statements of Non-use	0	1	0	1	0	0	10	2	7	21

KEY:

A — Teacher Constructed	F — District-Objectives Based
B — Teachers' Other Major Assignments	G — Standardized
C — Curriculum Embedded	H — Minimum Competency
D — School/Department/Grade Level	I — Statewide Assessment
E — Commercial Diagnostic	

what I'm telling them and trying to motivate them to do better than they've ever done before.

- You can't count on a score on one test too heavily. The kid could be sick or tired or just not feeling up to doing it that day. Maybe his parents had a fight the night before. Maybe he doesn't test well.

It seems, then, that part of what teachers "know" is that students can vary as test takers and that a variety of situational factors can influence students' test performance. Under these circumstances, teachers appear to reason, it is better to rely upon a variety of information sources — and especially on one's day-to-day experience with students in the variety of task and performance contexts that routinely recur in the classroom. If principals share this outlook, it may explain why they, too, routinely count on teacher's judgments, opinions, and recommendations (Tables 9 and 10 above).

CHAPTER 4

ADMINISTRATIVE LEADERSHIP: MONITORING AND SUPPORTING ASSESSMENT

A growing research literature demonstrates the importance of district and school leadership in the implementation and maintenance of particular education innovations, programs, and practices (e.g., Berman & McLaughlin, 1977; Bank & Williams, 1981b; Edmonds, 1979). In view of these findings, the Test Use in School Study sought to describe how, and how regularly, district and school administrators play leadership roles in local achievement assessment.

Exploratory fieldwork suggested that administrators' assessment related activities tend to fall into four general categories and to include both monitoring and supporting functions. The four categories include:

(1) *monitoring testing* — checking to see that appropriate assessment practices are followed.
(2) *linking test results with instruction* — reviewing test scores, examining their implications for instruction, communicating these to school staff, and monitoring instruction to assure that it attends to the areas that scores suggest should be emphasized.
(3) *providing staff development* — supporting assessment and test use by initiating in-service training and informational sessions.
(4) *facilitating routine classroom assessment* — initiating and maintaining technological and organizational arrangements that reduce teachers time on testing.

Fieldwork also indicated the range of ways in which district and school administrators commonly carry out each of these leadership roles. In addition, it confirmed that principals usually have much more reliable knowledge about their district's policies and practices than classroom teachers do.

CSE's national survey took these findings into account. Questionnaires examined the four types of activities listed above; specific questions and response choices were generally derived from the fieldwork. Questions about the role of district administrators were directed to principals, rather than teachers. Both principals and teachers were asked to report on certain school-level leadership activities.

The results of this inquiry are described and discussed below.

District Testing Programs Are Closely Monitored; Routine Classroom Assessment Is Not.

As Table 18 shows, most principals say that their district administrators closely monitor districtwide testing programs to be sure they are properly carried out. While fewer than half at both levels of schooling find that such oversight is regular or routine, many others note that it occurs "fairly often." Only 25% of the elementary principals responding and 32% of the in secondary principals report that their districts rarely or never check up on district testing.

In sharp contrast, there appears to be very little monitoring of routine classroom assessment. Administrators in most schools do not systematically review and critique the tests that their teachers construct. This practice is regular or frequent in only 13% of the elementary principals' schools and in 30% of the secondary principals'. (Administrative review of high-school final examinations, fieldwork suggests, may account for the difference in these percentages.)

Table 18. Monitoring Achievement Testing (Percentages of *Principals* Reporting the Regularity of Each Activity)*

	Elementary				Secondary			
	Routinely	Often	Rarely	Never	Routinely	Often	Rarely	Never
DISTRICT ADMINISTRATION								
Conducts observations and/or requires reports to see that all aspects of the district testing program are properly carried out	44	30	19	6	38	30	20	12
THE SCHOOL ADMINSTRATION								
Requires teachers to turn in copies of the tests they construct to be reviewed or critiqued	6	7	26	60	12	18	35	35
Requires that teachers turn in the scores or grades of the tests and/or assignments they routinely give in their classrooms (e.g., unit tests, chapter tests, etc.)	21	12	36	32	18	15	35	32

*Principals indicated the regularity of each activity from among the following response choices: 4 = happens regularly or routinely (i.e., on a systematic, periodic basis as part of routine procedure); 3 = is not regular or routine but happens fairly often; 2 = is not regular or routine and happens rarely; 1 = does not happen at all.

Monitoring of teachers' test results, it appears, is only slightly more common than the practice of reviewing their tests. A mere third of the principals at each level of schooling make it a routine or frequent requirement for teachers to turn in students' scores or grades on classroom tests and assignments. When they do so, furthermore, it may not be for oversight purposes. Fieldwork found one elementary school principal who did examine all the reading and math unit-test scores of each of his thirteen teachers' pupils in order to "keep track of how things are going and identify problems that should be discussed." Elsewhere however, principals gathered students' scores on commerical, curriculum-embedded tests on a *pro forma* basis and never examined them. They were used only to complete forms in compliance with evaluation requirements for a special program. In addition, several high school adminstrators mentioned collecting students' grades on final exams in case there are any complaints from parents about the course grades" or "in order to protect the teachers."

In summary, the results in Table 18 indicate that most school administrators do not check up very often on teachers' test designs, scoring procedures, or grade distributions. Rather, they appear to trust their teachers' professional competence in assessing student achievement. The next chapter offers further evidence to support this proposition. While few review teachers' assessment procedures often, over 80% of the principals studied express confidence that teachers construct tests of high quality (Table 25, page 60). All this is especially worthy of note given the importance generally accorded the results of teacher-made tests and assignments in a wide variety of school and classroom tasks (Tables 9 through 14 in Chapter 3).

Testing And Instruction Are Not Well Linked In Many Districts and Schools.

Evidence in the previous chapter (Tables 11 and 13) indicates that both principals and teachers tend to rely heavily on the results of many different types of tests as they go about planning curriculum and instruction. Nevertheless, it appears that a good many district and school leaders are doing less than they could to facilitate the use of test results in the planning and teaching process.

Tables 19 and 20 below list several very basic activities that district and school leaders can undertake toward linking test results with curriculum and instruction. As a first step (Table 19), districts can arrange testing and test scoring such that results are returned to schools at a time and in a format which permit them to be useful and used. Then, once the scores arrive in a school (Table 20), administrators there can initiate meetings with teachers to examine their implications: to identify and highlight the subjects and skills that seem to require greater (or less) teaching emphasis. If principals' perceptions are correct, however, these are consistent, routine procedures in only a minority of settings.

Over half (54%) of the high-school principals and nearly as many elementary-school administrators (47%) say that their districts rarely or never return test results in ways that make them useful for curriculum planning. Those who find that their districts do so regularly and systematically comprise only small proportions of the sample: 30% at the elementary level and 18% at the secondary level.

Most principals claim that they do better in reviewing and analyzing the test results with their teachers. Some 84% of those in elementary schools respond that they meet with teachers regularly or at least fairly often to discuss what test scores mean for instruction. Among the high school principals, 76% reply in the same way. But if their reports of district procedures for returning results are correct, many may be discussing scores that are outdated or otherwise inappropriate. Alternatively, principals may be using different standards to judge what is "routine" and "often" in describing their own behavior and their

Table 19. Linking Test Results with Instruction: District Leadership (Percentages of Principals Reporting the Regularity of Each Activity)*

	Elementary				Secondary			
	Routinely	Often	Rarely	Never	Routinely	Often	Rarely	Never
DISTRICT ADMINISTRATION . . .								
returns test scores in such a way that I can use them to decide on the skills and content we need to work on in our school	30	23	22	25	18	28	28	26
observes my work, reviews school plans, and/or requires written reports to be sure the school is emphasizing the skills or content areas that test scores show need emphasis in our school	32	34	23	11	26	29	38	7
establishes specific test score goals for our school to meet	20	16	19	46	19	19	30	32

*See footnote to Table 18 for a detailed description of these response choices.

Table 20. Linking Test Results with Instruction: School Leadership (Percentages of Principals and Teachers Reporting the Regularity of Each Activity)*

	Elementary				Secondary			
	Routinely	Often	Rarely	Never	Routinely	Often	Rarely	Never
DISTRICT ADMINISTRATION . . .								
meets with individual teachers, departments, and/or grade levels to review test scores in order to identify skills or content areas that need extra emphasis/less attention	34 (37)	48 (22)	17 (19)	1 (12)	25 (14)	51 (19)	21 (36)	3 (31)
observes teachers, reviews their plans, and/or requires written reports to be sure they are giving emphasis to the skills content, etc. that test scores show their students need to work on	53 (31)	30 (24)	17 (24)	0 (21)	40 (22)	40 (19)	17 (31)	2 (28)
considers students' test scores in evaluating teachers and/or establishes test score goals for teachers to meet	4 (6)	8 (8)	32 (15)	56 (70)	1 (12)	10 (5)	39 (10)	50 (72)

*Teachers' response are shown below principals in parentheses.
See footnote to Table 18 for a detailed description of these response choices.

districts'. Another possibility is that some principals, viewing the use of test data in instructional planning as a desirable practice, have exaggerated the frequency with which it occurs in their schools.

Teachers' observations (Table 20) support this last hypothesis. In general, they assert that meetings to link test information with instructional plans take place less regularly than principals maintain that they do. Assuming the salience of such meetings for teachers is the more important (since it is they, after all, who must put any instructional plans into effect), it appears that test-based planning occurs on a regular, periodic basis in about 37% of the elementary teachers' schools and 14% of the high-school teachers'. In another 22% of the former and 19% of the latter, it seems to occur fairly often. (Refer to the figures in parentheses in the first line of Table 20.) While these percentages are not insubstantial, they do suggest that many school leaders could be deriving greater value from their test scores than they currently are. In addition, many leaders at the district level could be doing more to facilitate this process by getting scores into principals' and teachers' hands in a timely and useful fashion.

Following through to be sure that test-based curricular and teaching plans are implemented is a next, fundamental step in linking testing with instruction. Thus, district administrators can visit schools, review their plans, and/or require written reports to be sure schools are emphasizing the skills or content areas that test scores show are in need of extra attention (Table 19). School administrators can take similar steps with classroom teachers (Table 20). Somewhat ironically, it appears that both district and school leaders pursue these monitoring procedures more regularly than they make test results and their implications accessible and clear to teachers. (Compare the first and second lines of Table 19 and Table 20. Once again, note the differences in principals' and teachers' reports in Table 20.)

As yet another step in holding their staff members accountable for test-based curricular and instructional plans, administrators can establish specific test-score goals for schools and teachers to meet. They can also take students' test results into account in teacher evaluation. Table 20 reveals, however, that these steps are rarely taken at the school level. Only 12% of the elementary-school principals and 11% of those in secondary schools say that they regularly or frequently set test-score goals for their teachers to meet or consider test results in teacher evaluation. As the next chapter demonstrates principals simply do not deem it appropriate to assess teachers, competence on the basis of their students' test performance. Most rely on their own observations of teachers work in the classroom for this purpose (Table 25, page 60).

Administrators at the district-level, on the other hand, are more likely to set test-score benchmarks for schools. Over all, 36% of the principals in elementary schools and 33% of those in high schools report that their districts do so routinely or often (Table 19.) This practice, survey results also suggest, occurs more commonly in districts serving lower socioeconomic groups than in those serving the well-to-do. Only 10% of the elementary and secondary principals in the highest socioeconomic districts sampled say that they routinely face district-established test-score goals. Among those in the lowest socioeconomic districts sampled, however, the figure is 40%.

Reviewing all the "routinely" and "often" columns in Tables 19 and 20, it is evident that roughly a half to two-thirds of the principals' districts and schools manifest some concern that test scores be used in curricular planning and instruction. Nevertheless, it is also apparent that comparatively few administrators *routinely* take steps to be sure that test scores are readily accessible or *routinely* review those test scores with their faculty members. More, but still relatively small percentages of administrators, routinely check to see that test-score-based curricular and instructional decisions are actually carried out in classrooms. Even fewer choose to hold schools and teachers accountable for such decisions by projecting test-score objectives for them to achieve. Considering test results in evaluating teachers, moreover, is generally avoided. All of this — plus certain apparent inconsistencies in principals' reports and the divergence of teachers' and principals' — suggests that in most districts and schools the links between testing and instruction are very loose indeed, especially at the secondary level. Fieldwork during the Test Use in Schools Study supports this finding, as does on-site research conducted in other CSE projects (e.g., Bank & Williams, 1981a).

Teachers Average Seven to Eight Hours a Year In Assessment Inservice; Explanations of How To Administer Tests and of Test Results Are the Most Common Topics.

Studies have repeatedly revealed that teachers receive little preservice training in testing and measurement (e.g., Coffman, 1983; Yeh, 1978). This is one reason why their inservice activities in assessment are of special interest. What is more, it appears that staff development is a critical factor in districts' establishment of systems to link testing-evaluating instruction linkage systems (Bank & Williams, 1981a). Districts' and schools' staff development and informational activities in the area of assessing student achievement assessment, therefore, were given considerable attention in the CSE national survey.

Principals' responses show that district-sponsored staff development in assessment occurs routinely or often in 61% of their elementary schools and 57% of their high schools. School-supported inservice takes place, they collectively report, only slightly less regularly (Table 21). Allowing teachers extra pay or time away from the classroom to help develop tests and related materials appears to be a somewhat less widespread practice. Some 41% of the elementary and secondary principals say that it happens routinely or frequently in their districts.

These figures suggest that most districts and schools give considerable attention to training teachers in assessment and to a lesser degree, utilize teachers' skills in local test development. Once again, however, teachers' reports present a more modest picture. The elementary teachers surveyed estimate that they had spent, on the average only six hours in district or school-suported inservice training on student assessment during "the last two years." Secondary teachers judge that they had spent an average of only five hours thus engaged in the same period. During those two years, meetings to select tests, to construct them, or to help formulate testing policies consumed another eight hours for elementary teachers and an additional eleven for high-school instructors. (See Table 22.) All told, then, it appears that teachers average about seven or eight hours a year on all district- and school-sponsored inservice activities connected with assessment. Of this total, teachers spend about two-and-a-half or three hours expanding their assessment skills.

These estimates should be taken as extremely rough, based as they are on teachers' recollections over two years. They do, however, put principals' estimates of district and school support in perspective. If local educational agencies are devoting a great deal of time to developing or employing teachers' assessment skills, that time is not particularly salient for most teachers.

Table 21. Supporting Assessment Through Staff Development and Release Time (Percentages of Principals Reporting the Regularity of Each Activity)*

	Elementary				Secondary			
	Routinely	**Often**	**Rarely**	**Never**	**Routinely**	**Often**	**Rarely**	**Never**
DISTRICT ADMINISTRATION provides speakers, workshops, printed material, etc. in an effort to help teachers expand and update their skills and understanding in the area of student assessment	26	35	26	13	22	35	32	11
provides released time and/or extra pay for teachers to help develop tests (or curriculum materials *that include* tests)	13	28	25	34	12	29	33	26
THE *SCHOOL ADMINISTRATION* brings in speakers, workshops, printed material, etc. to help teachers update and further develop their skills and understanding in the area of student assessment	22	32	36	10	9	35	47	9

*See footnotes to Table 18 for a detailed description of these response choices.

Table 22. Teachers' District and School In-Service Time on Assessment (Reported in Average Number of Hours Spent Over the Last Two Years)*

	Elementary Teachers	Secondary Teachers
Meetings within my district or school to select or construct tests and/or to help establish testing policy	8	11
District or school supported inservice training on topics related to student assessment (testing, other techniques)	6	5

*The figures given here are rounded to the nearest hour. They are based on teachers' responses to the following direction: "For each activity below in which you have participated, indicate the approximate TOTAL number of HOURS you spent in the last two years."

Table 23 elaborates on these findings, showing how teachers spend their staff development time. For the most part, they attend explanations of state, district, or school test results; receive directions on how to administer required tests. Inservice training that would help teachers develop or expand classroom assessment skills, the table shows, tends to occur far less frequently. Thus, for instance only about a fifth of the teachers in each category report receiving instruction in "how to construct or select good tests." Information on alternatives to testing is provided just as rarely for secondary teachers, although some 54% of the elementary teachers do report staff development on this topic. Training in the use of test results to improve instruction is evidently provided for 35% of the elementary teachers and about 20% of the secondary teachers sampled.

Two other staff development activities on the list can be construed as aimed directly at improving students' test results, "How to tie what is taught more closely to the skills, content covered on required tests" and "Presentation of published materials designed to prepare students for particular tests or to improve test-taking skills." From a quarter to a third of the secondary teachers and 40% to 50% of elementary teachers have received training in these areas.

In summary, it appears that districts and schools are doing much less than they could to build teachers' competencies in achievement assessment. This is especially true for high-school teachers.

Table 23. Teachers' Participation in Staff Development (Percentages of Teachers Who Report Joining in At Least One Session on Various Topics During "the last two years")

Topic	Elementary	Secondary English	Secondary Math
(1) Analysis and explanation of state, district, or school test results	84	70	60
(2) How to administer tests required by my state, district, and/or school (procedures to follow, etc.)	78	54	46
(3) How to interpret and use results of different types of tests (e.g., norm-referenced and criterion-referenced tests and their applications)	59	35	34
(4) Alternative ways (other than tests) to assess student achievement	54	25	21
(5) How to tie what is taught more closely to the skills, content covered on required tests	50	37	25
(6) Presentation of published materials designed to prepare students for particular tests or to improve test-taking skills	41	32	29
(7) Training in the use of test results to improve instruction	35	21	19
(8) How to construct or select good tests	20	23	18

Resources To Facilitate Routine Classroom Assessment Are Not Widely Available; But Where They Are Available, They Are Used.

Survey and fieldwork results discussed in Chapter 2 demonstrate that teachers spend considerable time constructing, grading, and recording the results of their own tests and assignments. Administrators can help teachers reduce this time by initiating and supporting technological and organizational arrangements that facilitate their testing work. Among those that fieldwork found to be available were banks of test items, computerized test scoring and analysis and, of course, paid paraprofessionals or volunteers to assist teachers in reading and grading tests and assignments. In addition, fieldwork suggested that some principals provide special time and support for teachers to develop tests that they can use in common with classes in the same grade level, subject, etc.

While fieldwork and questionnaire piloting indicated that this was a reasonable list of resources to investigate in the national survey, survey reports show that three of the four are unavailable to large proportions of survey respondents (See Table 24). The exception, of course, is "other teachers with whom I plan and develop tests or other evaluation assignments," but only about a quarter of the elementary-school teachers and a similar fraction of the secondary-school teachers report taking advantage of this resource at least monthly. Some 45% of the secondary teachers say that they construct tests with others a few times a year, and fieldwork suggests that this often occurs as teachers in the same department conjointly devise mid-term and final exams.

Table 24. Available Resources for Testing (Percentages of Teachers Reporting)

Resource	NOT AVAILABLE	AVAILABLE		
		Not Used	Used Once To Several Times/Year	Used at Least Once/Month
Item banks of test questions upon which I draw in	71	4	8	16 Elementary
making up my tests.	51	8	24	16 Secondary
Other teachers with whom I plan and develop tests or other	37	12	26	24 Elementary
evaluation assignments.	21	10	45	24 Secondary
Someone who helps me read, grade, or correct	69	6	4	21 Elementary
tests and assignments.	70	5	4	21 Secondary
Quick, computerized scoring and analysis	64	2	30	4 Elementary
of tests.	58	16	22	4 Secondary

Computerized test scoring and analysis is used a few times annually by a quarter to a third of both the elementary and secondary teachers sampled. Fieldwork indicates that this probably reflects the use of small, on-site optical scanning machines for scoring multiple-choice and similar "objective" tests. The number of districts and schools with more sophisticated equipment that analyze students' errors is still quite small. Some districts, however, have developed computer programs for scoring unit and chapter tests and simultaneously analyzing individual students' strengths and weaknesses on the skills they cover.

A final point: in general, nearly all those teachers who have access to the resources listed indicate that they use them at least sometime during the school year.

CHAPTER 5

PRINCIPAL'S AND TEACHER'S PERCEPTIONS AND BELIEFS ABOUT TESTING

Previous chapters have focused on what teachers and principals report that they do in assessing students' achievement, in using assessment results, and in monitoring and supporting assessment. Here, attention shifts from what teachers and principals do in assessment to what they perceive, believe, and value as they do it.

Three complementary objectives shaped CSE's exploration of principals' and teachers' viewpoints on testing. One was to elaborate and clarify, confirm or disconfirm the values and beliefs suggested by principals' and teachers' assessment practices. A second objective was to gather their perceptions of current testing trends and policies and of how these are affecting the schools. In the widespread debate over testing and its uses, administrators and teachers in the schools have had little direct voice. Here was an opportunity to solicit their views. A third objective was to examine relationships between assessment attitudes and activities: to learn whether certain sets of beliefs seem to co-occur with and "explain" certain practices or, on the other hand, whether particular practices (in staff development, for example) seem to coincide with and account for particular sets of beliefs. Such relationships could point the way toward policy and action in local school districts and schools.

Toward these ends, the survey questionnaires presented principals and teachers with sixteen statements and asked them to indicate strong agreement or agreement, disagreement or strong disagreement with each. The statements for principals and those for teachers varied slightly in phrasing, taking into account differences in their respective roles. Nevertheless, both forms of the questionnaire covered identical topics: (1) the quality of achievement tests; (2) their value or usefulness; (3) effects of testing on the school; (4) the fairness and desirability of minimum competency (proficiency) testing; (5) educators' accountability for students' test results; and (6) the importance of testing as a local educational issue.

Respondents' perceptions and beliefs regarding the first four issues evolved as especially relevant in later analyses. They are emphasized in the discussion below; their relationships with other study findings are described in the next chapter. Viewpoints on issues (5) and (6) are mentioned briefly in this one. As in previous sections, information from fieldwork interviews serves to supplement and elaborate the survey results.

Principals: A Pro-Testing Perspective

Testing appears to be a central issue in the professional lives of most of the principals studied. Nearly two thirds report that it receives "a good deal" of discussion in their districts. What is more, a substantial majority seem to approach their discussions with a highly favorable view of tests and testing. (Refer to Table 25.)

Principals judge that the quality of tests is generally high. Eighty percent or more of those who responded apply this judgment to tests that accompany published curriculum materials, to tests developed by their districts, and to the tests constructed by the teachers in their schools. A similar proportion (82%) concludes that standardized tests are fair for most students.

Unfortunately neither the survey nor project fieldwork was able to explore exactly how principals arrive at these judgments. Principals' broad confidence in test quality, however, is worthy of note in itself. It can help to explain their regular use of test results in a variety of routine tasks (Tables 9 and 10, pages 31 and 32), as well as their general belief in testing's validity and value (discussed next). Later, as the policy implications of this study are examined, principals' confidence in test quality will be cited again.

Most principals see testing as valid and valuable. Principals, we have seen, rely on test scores most heavily for planning curriculum and (especially) for reporting school achievement to district officials, parents, and the general public. These uses can follow from district directives, public expectations, and other forces beyond principals' control. Be that as it may, most principals seem comfortable using test results in these ways. On the whole, they believe test scores accurately reflect their schools' performance, and they generally see testing as an asset.

By an overwhelming majority, principals reject the view that schools should not be held accountable for their students' test results. (See Table 25, "Accountability."). They appear to accept that it is what goes on in school — and not, for instance, students' native abilities, their parents' support, or the community environment — that is primarily responsible for how students do on tests.

In a consistent set of responses, two thirds of the elementary-school principals and three quarters of those in high scools find that test scores provide "a fairly good index of how a school is doing." As one California high-school principal explained in an interview:

> I'm not a believer that test scores tell all. Many factors contribute to outcomes and they're not all revealed in test scores. But they are important indices. They're something we should take a look at among other data . . . Like with our [standardized test and state assessment] results, I keep a running tally of the means and of where we are, so that I'm aware of the progress and of where our students may have had some difficulty. And we share that with the math and English departments, particularly, and with the rest of the staff.

At an Iowa high school, the principal volunteered a similar perspective:

> I don't know that test results *per se* would change specific instruction much, but if year after year after year we had a department rating low, we would certainly look at several things. We'd want to talk to the people [in that department] to see what the problem is.

These remarks reflect a qualified, or cautious, acceptance of test scores as "indices" of school performance. Fieldwork suggests that such a stance is common among both elementary- and high-school administrators: It may well underlie their questionnaire response.

While most principals maintain that test results reflect overall school performance, many fewer believe that individual teachers can be held accountable for them. Only 32% of the elementary-school principals conclude that "test results can be used to evaluate teachers' effectiveness or competence." Among the high-school principals responding, 49% agree. Recall, however, that principals at both levels claim that they in fact place little emphasis on test results in teacher evaluation. In general, they tend to trust their own observations of their staff's teaching skills. (Again, refer to Table 9, page 31.) In some cases, of course, administrators who would use test scores to evaluate teachers literally cannot do so. As a result of district policy or an agreement with teachers' representatives, they never receive classroom-by-classroom breakdowns of students' test results. But many seem to concur with the views of an elementary-school principal who argued:

> You can't evaluate teachers from the office. You need to be in the classroom and be there frequently. Low [test] scores could mean we're

Table 25. Principals' Views on Testing and Related Issues (N = 221)

Issues and Items	Percentage of Principals in Agreement	
	Elementary	Secondary
Testing As A Local Issue		
Testing is an issue that is discussed a great deal in our district	61	67
Quality of Tests		
The quality of tests that come with published curriculum materials is generally high	86	88
The quality of our district-developed tests is is generally good	84	86
The teachers in my school develop tests of high quality	79	85
Standardized tests are fair for most students	82	82
Value, Usefulness of Testing		
Test scores are a fairly good index of how well a school is doing	68	74
Student test scores can be used to evaluate teachers' effectiveness or competence	32	49
The pressure that required testing exerts upon me and the teacher in my school has a generally beneficial effect	62	62
As a result of minimum competency testing (and similar programs), parents are contacting the school . . . more frequently or in greater numbers	56	54
Desirability/Fairness of Proficiency Testing		
Minimum competency/proficiency tests should be required of *all* students for promotion at certain grade levels and for high school graduation	58	70

Minimum competency/proficiency/functional literacy tests are generally fair for all students	58	72
Effects on the School		
In the last five years, the amount of testing required by our district, state or federal program(s) has increased dramatically	68	75
As a result of testing programs (for minimum competency, etc.), more time is being spent on reading/English and math instruction in our school	71	76
The amount of time that is given to required testing and the preparation for it in my school is too great	31	26
Accountability For Test Results		
Schools should not be held accountable for their students' scores on required or standardized achievement tests	37	30
Schools should not be held accountable for their students' scores on minimum competency/proficiency/functional literacy tests	30	21

> not providing the supplies and materials. They could mean working conditions are a problem. It could be the types of students they're getting. It could be me. There are too many factors to say, "the scores are low, therefore the teacher is ineffective."

This way of thinking emphasizes that it is the school as a whole — and not the individual classroom teacher — that produces test results.

For many principals the value of testing extends beyond scores and their uses to the influence testing has on the school community. Among respondents at both levels of schooling, 62% find that testing requirements exert a beneficial pressure on their teachers and on them. This lends support to those contemporary school reformers who suggest that stiffer testing requirements will help raise educational standards.

At least one type of testing requirement seems to influence many parents' behavior. In most states, laws creating minimum competency (proficiency) testing also specify that parents be informed of their children's results. Districts and schools routinely encourage parents to discuss these results with school officials, and some schedule conferences with parents whose children have failed to meet minimum standards. A majority of principals responding (about 55%) observe that these measures have stimulated greater contact between parents and schools. Where program requirements are more stringent, i.e., where proficiency tests must be passed for promotion to certain grades and/or for high-school graduation, the proportion of principals who note increased parent contact is somewhat greater (slightly over 60%).

Principals favor proficiency testing for promotion and graduation. Some 70% of the study's high-school principals advocate that students should be required to pass a minimum competency or proficiency test for promotion at certain grade levels and for high-school graduation. A similar proportion (72%) finds that tests of this type "are generally fair for all students." Principals of elementary schools tend to support both views, but by a smaller majority (58%). Principals' opinions on these issues did not vary substantially according to the requirements now in place in their states and districts.

Here, it is worth noting that CSE data (Choppin et al., 1981) show 20% of the nation's school districts, serving roughly 35% of its pupils, require proficiency tests for promotion to certain grades and/or for high-school graduation. Another 35% of the districts, with about 32% of the nation's students, also work under state minimum competency/proficiency mandates. Here, however, the tests are used only for diagnostic purposes, not as promotion or graduation prerequisites. The remaining districts, with 34% of the nation's school enrollment, operate without state-mandated minimum competency/proficiency testing.

although a few of these have established their own proficiency requirements. State laws have been in flux and the figures may have changed somewhat since these data were collected. Nevertheless, the picture outlined here should help to put principals' viewpoints in perspective.

Principals find that more required testing has led to more basic skills in the curriculum. For 68% of the elementary-school principals and 75% of those in high schools, the amount of testing required by their district, by their state, or by federal programs has increased dramatically "in the last five years" (1977-1982). Simultaneously, nearly three quarters find that, as a direct result of testing programs, more instructional time is being spent in their schools on the basic-skill subjects of reading/English and mathematics. Principals' responses on these two issues, furthermore, are related at a statistically significant level; they tend to be consistent much more often than not. (See Table 26.) All this suggests that if most principal's perceptions are accurate, a recent, marked increase in the amount of required testing has had a discernible impact on the curriculum: it has pushed instruction toward the basic-skills subjects that required tests emphasize and (probably) reduced the teaching-learning time available for other subjects. For the most part, however, principals do not find testing requirements troublesome. Fewer than a third say that their schools spend two much time on required testing and the preparations for it. (See Table 25.) This seems in line with the majority belief that testing exerts a positive influence on the schools.

Teachers: Qualified Support For Tests and Testing

As teachers received their CSE questionnaires in the early 1980's, social problems such as classroom discipline, school safety, and students' drug and alcohol abuse captured medical attention and preoccupied many educators. Even compared to such problems, however, teachers in a majority of schools could define testing as an important concern (Table 27), just as principals in a majority of districts do.

More broadly, teachers' responses reflect greater concern about tests, testing, and their effects on schools than do principals'. Teachers do generally support testing, but that support is less consistent, less overwhelming numerically, and (thus) more qualified than the support that principals express. (Refer to Table 27 here and throughout.)

Most teachers agree that test quality is high; although by narrower majorities than principals, well over 70% of the teachers responding have decided that the content or skills covered by required tests, whatever their type, is similar to the material that they actually teach. Most

Table 26. Relationship Between Principals' Responses: Increase in Required Testing and More Time on Basic Skills

		Testing Has Led To More Instructional Time On The Basic Skills		
		Agree	Disagree	
Required Testing has Increased Dramatically	Agree	114	34	148
	Disagree	36	21	57
		150	55	

$x^2 = 37.83, p < .001$

(60% — 62%) also agree that the tests developed in their districts are "very good." Opinion on the quality of commercial tests tends to divide by grade level. Some 59% of the elementary-school teachers find commercial tests (such as those that accompany reading and math series) "are usually of high quality," but only 46% of the high-school teachers concur.

Teachers seek tests that they find fair and useful. It is impossible to know, of course, exactly what criteria the survey respondents use to assess test quality. Other aspects of CSE's Test Use in Schools Study, however, provide some clues: they suggest that teachers are most concerned about the fairness and practical utility of tests.

Results of an earlier CSE questionnaire study of testing in five California school districts (Yeh, 1978) were reanalyzed in planning for the national survey under discussion here. Among the 256 elementary-school teachers who responded, three criteria stand out as most important in selecting tests. Listed in descending order of importance, they are (1) the similarity of test material to what is presented in class; (2) clarity of test format; and (3) the ease with which the test can be administered and/or scored. The first two criteria reflect teachers' interest in test fairness; the third, their desire for practical utility.

Concern with these same three features recurs throughout teachers' interview comments on test quality. In addition, interviewee's remarks reveal a fourth consideration, another dimension of tests' practical utility: the degree to which tests yield information that teachers can in fact use in their routine teaching tasks. The words of one fourth-grade instructor epitomize this concern:

> I don't feel we need to test, test, test; but if the information is something I can use to prescribe instruction, I really don't mind giving it.

These criteria provide insights into teachers' views of test quality and into their test-use practices.

Teachers in both elementary and high schools tend to count the results of their own, self-constructed tests as especially important for routine instructional tasks (Tables 12 and 13, pages 36 and 37). Asking teachers to rate the quality of their own tests seemed unnecessary, but note that they do have, from the teacher's perspective, all the qualities of good assessment instruments. In making their own tests, teachers can suit themselves regarding the fit between what is tested and what is taught.

They design the format. They determine how easily the test can be administered and scored. They also control the timing of the test, when the results become available, and other factors that allow the measure to serve their everyday, practical needs.

In interviews, teachers at the elementary level regularly associate these same qualities with the commercial tests with which they work most frequently — those that accompany their basal reading and mathematics texts. As one explained:

> The district tells us we have to use the tests that go with the book — the ones you buy from the publisher. But we'd all use them anyway. They match with the skills we're teaching and present things the same way [that the book does], so they're really convenient.

This widespread view can help to explain why the majority of elementary-grade respondents rate commercial tests as high quality, as well as why most rely heavily on the results of commercial, curriculum-embedded measures (Table 12, page 36).

High school teachers mention these same criteria in discussing commercial tests, but they speak of these tests more negatively. With greater latitude in selecting their course content, they frequently find commercial tests less useful than their counterparts in the lower grades.

Table 27. Teachers' Views on Testing and Related Issues

(Elementary Teachers: N = 486)
(Secondary Teachers: N = 385)

Issues and Items	Percentages of Teachers in agreement		
	Elementary	Secondary	
		English	Math
Testing As A Local Issue			
In our school, testing programs are generally held to be much less important than the social problems with which we are concerned	39	32	42
Quality of Tests			
Commercial tests are usually of high quality	59	46	46
The tests developed in our district are very good	62	62	60
The content (or skills) on most required tests is very similar to the content (or skills) that I teach	72	77	79
Value, Usefulness of Testing			
Testing motivates my students to study harder	73	80	93
The pressure that testing exerts on the schools has a generally beneficial effect	48	60	72
As a result of minimum competency testing (and similar programs) parents are contacting the school . . . more frequently or in greater numbers	53	42	36

Desirability, Fairness of Proficiency Testing			
Tests of minimum competency/proficiency should be required of all students for promotion at certain grade levels and for high school graduation	81	86	90
Tests of minimum competency/proficiency are frequently unfair to particular students	58	48	35
Effects on the School			
Recently, I have been spending more teaching time preparing my students to take required tests	46	41	30
Tests of minimum competency have affected (would affect) the amount of time I can spend teaching subjects or skills that the tests do not cover	62	62	42
Basic skills teaching (including remedial work) is now consuming a substantially increased proportion of our school's educational resources	88	84	74
The proportion of our school's resources now allocated to basic skills teaching is so great as to detract from the quality of our overall educational program	23	28	21
Accountability For Test Results			
Teachers should not be held accountable for students' scores on standardized achievement tests or tests of minimum competency	71	61	61

An instructor of senior English spoke for many of his colleagues in saying:

> I'll occasionally use a [curriculum] kit or package as is, and then if there's a test that comes with it, I'll use it. But in most units I'm putting together materials, combining things from [many sources]. The only test that will cover it all is the one I make up myself.

The remarks of a geometry teacher pinpoint another limitation of commercial tests:

> We rely fairly heavily on the unit post tests we developed as a department . . . We don't use the book tests. Every one of our courses has performance objectives, and we have designed each unit test to validate to the performance objectives for the course. The book tests just don't do that . . . Our biggest concern is the validity factor, in terms of our objectives for the course.

It is, perhaps, for reasons such as these that 54% of the secondary English and math teachers do not consider commercial tests "of high quality." Such views can also help illuminate why high-school students spend 75% of their total testing time taking teacher-made tests (Table 4, page 17).

The broad popularity of district-developed tests (60% — 62% rate them "very good") can also be traced to their fairness, or validity, and practical utility.

> That computer-processed data [on our district's objectives-based unit tests] can really be used with those kids that need help. It does a better job [than the other tests available] of identifying students and students' needs. . . I can work on objectives 2, 3, 5 and 9.
>
> The district [testing] system is important because it's the only thing you can pass on to other schools which is meaningful to everbody. There's a lot of movement in this town, and the elementary schools, many of them, use different [text] series.

When district-made tests fail to meet these criteria, however they can be ignored or deemed a burden.

> You've already tested your kids with the test that comes with the series. Then you have to give the district tests, 'cause they require you check off the skills on the [record-keeping] card when they complete them. But the district test doesn't really fit with the way our series lays things out, so it's a waste — just more red tape.

> No one uses the [district-constructed] unit reading tests anymore. We need to, before we adopted the new series a couple of years ago. But now they aren't really valid.

A sizeable minority of teachers does not find district-developed tests "very good"; problems such as these may explain their judgments.

Finally, a word or two about teachers' views of required tests is appropriate here. Most survey respondents agree that these measures generally cover what they teach (Table 27), but many fewer count their scores as of great importance (Tables 12 and 13, pages 36 and 37). Interviews offer an explanation for this apparent discrepancy: standardized and other required tests often fail to meet practical utility criteria.

> The [standardized test required anually in our district] is almost useless in the spring, which is too bad, because I feel there is some valuable information there, progress and growth. But we get the scores the last week of school.

A high school teacher added:

> You don't get individual students' scores on the [state-assessment test], and the standardized results, they're there in the [cumulative-record] folders. But I have 150 students. I don't have time to go down to the office and look through all those folders.

More generally, nearly every teacher interviewed echoed views of an elementary-school teacher in urban New England:

> I think that the children feel good about [a test] and I feel good about it if I can see where it is actually helping the child and you can put it in context. But when you pull it out of the context, out of the classroom teaching situation and the actual curriculum, and give a child a test just to rate him nationwide or whatever, that bugs me. It really bothers me.

This statement summarizes teachers' interest in tests that cover what they believe they are teaching and also provide information that teachers can use in their routine teaching tasks.

Teachers value testing as a motivator. Nearly three quarters of the elementary teachers and even larger proportions of the secondary instructors (Table 27) claim that testing motivates their students to study harder. This can be a primary reason for some classroom assessment. As one high-school English teacher explained in her interview:

> I'd like to eliminate the quizzes that I give every week or so, but I have to do it to motivate the students to do the reading.

Most high-school teachers (60% in English; 72% in mathematics) also concur that the pressure that testing exerts on the schools has a generally beneficial effect. "It's kind of nice to get results back," said one who was interviewed. "It does give you more of a feeling of accountability and it's not overwhelming." Another added:

> I think that within this city there has been a lack of standardized testing, which I think has allowed things to go downhill. That is, if you don't measure versus some outside standard you don't know how good or bad things are going in the system, and it can just tend to get worse.

At the elementary level, however, fewer teachers (48%) agree that the pressure generated by testing is beneficial. One sixth-grade instructor voiced a concern felt by many others who were interviewed:

> There's too big a trend to judge teachers and schools by tests. They publish test results in the papers, and people use them to judge teachers and rank schools. This is the danger, [of testing], using the results in the wrong way.

Indeed, most teachers who responded to the survey (but somewhat fewer at the secondary level) assert that teachers should not be held accountable for students' scores on standardized or minimum comptency tests. (See Table 27, "Accountability for Test Results.") It appears, then, that many teachers (along with their principals) believe that schools, but not individual faculty members, bear responsibility for how learners perform on achievement tests.

About the same proportion of elementary-grade teachers (53%) as principals (56%) observe that parent-school contacts have increased as a result of minimum competency testing and similar programs. Only a minority of high-school teachers agree: 42% in English and 36% in math; as compared to 54% of their principals. It may be that parents speak more frequently with central office personnel than with teachers about their high-school students' scores. It may also be, as many teachers argue, that parents' active involvement with their children's schools diminishes as their youngsters proceed through the grades. Whichever the case, some teachers of secondary schools fault parents for their lack of concern. An English Department chairperson captured the feelings of many when he reported with frustration that:

> The point was, the legislature wanted to test [for minimum competency] and to assure effective communication, with the possibility of remediation, before the kid goes out [of high school]. . . . We had a form letter we sent out to about 150 parents where the students failed and couldn't

> graduate unless they got it together and passed. It said something like, "Your child has failed the following competencies" — there was a place to check which ones — "and we'd like you to come in and discuss this." Well, out of 150 parents only six, I think it was, actually showed up.

In summary, then, most teachers believe that testing exerts useful pressure on students, but their opinions are more divided about testings' effects on educators and parents.

Teachers heavily favor proficiency tests as promotion and graduation requirements, but many doubt that such tests are uniformly fair. From 80% and 90% of the survey respondents (Table 27) believe that all students should be required to pass proficiency tests in order to win promotion to certain grades and to graduate from high school. Interviewees' arguments in support of this position were usually quite general. "It's good for the student to know that he has to pass a certain level of competency," said one. Another simply asserted, "Students who are incompetent should be failed "At the same time, a majority of elementary-school teachers (58%) and substantial proportions of high-school instructors (48% in English; 35% in mathematics) judge that minimum competency (proficiency) tests "are frequently unfair to particular students."

Holding both these views simultaneously, as many teachers obviously do, does not necessarily signal inconsistency or an indifference to fairness. One can support the general concept of minimum competency requirements while doubting the uniform fairness of the particular tests now in use. In fact, there is evidence that as teachers actually experience minimum competency testing for promotion or graduation, they become more concerned about the fairness of the tests, more cautious about using them as gatekeeping standards, or both. This is exactly what Table 28, below, demonstrates. (Compare teachers' combined, mean responses on the fairness and should-be-required-for-promotion/graduation statements. Those of teachers in states where such requirements are now in effect are significantly lower — significantly less "pro-competency testing" — than those of teachers elsewhere.)

Fieldwork interviews reveal some of the kinds of experiences that can lead teachers toward more circumspect views of the fairness and desirability of testing for promotion and graduation.

> I wanted to tell you about the competency tests [said one high-school English teacher in a state that requires them for promotion and graduation]. I'm not happy with them, although I was on the committee that developed them for our district. There are eight competencies the [high

> school] kids have to pass. . .in one, they have to read a bus, train or plane schedule and answer eight questions about it. When we gave the bus schedule, we found that the black kids, the Hispanic kids — they ride the bus more and they did distinctly better on that than your more suburban kids, the white kids. Kids here at this school and others from, well where they're more likely to take the bus, they had better results. There's clearly cultural bias here. . . Another competency is filling out a job application, a standard form. [He shows one]. See, now if the student goes over the the line here as he fills this in, that's counted as an error. So some of this is very trivial, unfair really. . . There are other problems, too, and it's difficult figuring out how to resolve them. You begin to question whether you can ever come up with a test that's really fair.

Another teacher of high-school English cited inequities in how his district handles minimum competency requirements:

> The value of the district competency tests is that they are very explicit. Nobody has any questions about what's being tested. . . And I believe in failing a student for being incompetent. But you have to place responsibility on the students to work their way through [the tested skills] step by step. Here, a sophomore can pass part of the English [competency] requirement, fail others, and be passed right through all of his other classes and not be able to write a decent letter, not be able to demonstrate eighth-grade skills. So now, as a senior, they have special tutoring on how to pass the test and they graduate as a competent senior. That's not fair to anyone, either the kid who goes that route or the one who really masters the skills.

Thus, while there is among teachers a general enthusiasm for minimum competency tests as requirements for promotion and graduation, there is also notable concern about the fairness of these tests. This concern is significantly greater, and questions about the requirements themselves loom larger, where teachers have had to operate under testing-for-promotion/ graduation mandates.

Most teachers find an increased curricular emphasis on basic skills, due at least in part to testing, to be acceptable. As reported earlier, the vast majority of principals have noted a dramatic increase in required testing through recent years. Such testing — usually in the form of standardized batteries other minimum competency measures, and state assessment instruments — typically places heavier emphasis on basic reading, English, and mathematics skills than it places on other areas of the curriculum. Citing this fact, critics frequently argue that burgeoning testing requirements are "contracting" public school's curricula: forcing them toward a focus on basic skills at the expense of other subjects.

Table 28. Teachers' Views on the Fairness and Desirability of Minimum Competency Testing (MCT), By Current State Requirements*

State Requirement	SECONDARY[1]	ELEMENTARY[2]
MCT required for promotion/graduation, state-mandated measure	3.56	4.24
MCT required for promotion/graduation, local choice of measure	3.76	4.29
MCT required for diagnosis, state-mandated measure	3.93	4.38
MCT required for diagnosis, local choice of measure	4.20	4.96
No MCT required	4.16	4.79

**Explanation.* The values on this scale range from 2 (a strongly *negative* view of MCT) to 8 (a strongly *positive* view of MCT).

The scale shows the mean (or average) combined responses of teachers in each category to two survey statements:

(a) "Tests of minimum competency/proficiency are frequently unfair to particular students"; (1 = strongly agree, 2 = agree, 3 = disagree, 4 = strongly disagree);
(b) "Tests of minimum competency/proficiency should be required of all students . . . for promotion . . . and for high school graduation"; (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree).

1. Differences between groups statistically significant at $p < .05$
2. Differences between groups significant at $p < .01$

Principals concede that testing programs have caused more instructional time to be spent on basic skills instruction, but there is nothing to suggest that they find this troubling. (Table 26, page 64).

On the whole, teachers appear to support their principals and to reject the critics' argument. Along with the school administrators who responded, the teachers surveyed report a marked increase in basic skills instruction. Some 88% at the elementary levels, 84% in high-school English, and 74% in high-school mathematics agree that "basic skills teaching. . .is now consuming a substantially increased proportion of our school's educational resources." Only about 25%, however, feel that this detracts "from the quality of our overall educational program." (See Table 27.) Furthermore, fewer than half the teachers surveyed say that they have spent more time recently preparing their students for required tests. (At the elementary level, 46%, in secondary English and mathematics, 41% and 30%, respectively).

The "testing contracts the curriculum" argument does draw some support in survey responses, however. Teachers who find they are devoting more teaching time to preparing learners for required tests constitute a sizeable minority, as the figures just cited indicate. Representing their views, one teacher of grades 3 and 4 said,

> I'd like to cut all the testing down to about half. It seems like everything is testing; everything is evaluating. It is so continuing, it's almost suffocating. We have no time for any music or art. My kids used to learn English through writing stories and newspapers. We have no time for any of that. This is just cut-and-dry teaching, drill on tested skills.

In addition, a great many teachers believe that minimum competency mandates have affected (or would affect, if instituted) the amount of time that they can spend teaching skills and subjects not covered by these tests (62% in the elementary grades; 62% in high-school English; 42% in high-school math). Some of the teachers interviewed during fieldwork explained how this can happen. Discussing a math competency measure her students had to take, a fifth-grade teacher remarked,

> Ahead of time, because the format of the test is so different [from the tests my students usually take], we had to have the kids do worksheets and so on of that type so that when they did take the test, they were familiar with how to go about it, the mechanics of the test. Now, that's all time out of the classroom, and I couldn't use the scores for a thing.

A high-school instructor in a course called Consumer Math added:

> Well, see they use this course for kids who have failed the [proficiency] tests. So what I do, I spend the first four weeks doing nothing but reviewing the skills and having them take old versions of the test, the first month of school, really. Then you see which kids are going to have trouble on which of the four tests, then that's what you teach them.

Still another explanation of minimum competency testing's influence on the curriculum was offered by an algebra teacher:

> The first time they gave [the state proficiency test, required for diagnostic purposes only], I found there were kids having problems with certain things, and we really didn't emphasize those too much. So I went back and taught those things, which meant I dropped other units we'd usually cover.

All in all, however, most teachers appear comfortable with the increased emphasis on basic skills that they find. And while most believe that minimum competency requirements affect what they teach, only a minority conclude that they must spend more time preparing students for required testing.

Where districtwide socioeconomic status (SES) is lower, teachers find more emphasis on tested and basic skills. Individual teachers' responses on the four survey statements just discussed — those listed under "Effects on the School" in Table 27 — tend to correlate highly with one another. It is reasonable, then, to sum their responses on these items to obtain an aggregate indicator of the perceived emphasis on tested and basic skills. CSE survey analysts did so in an effort to determine whether this emphasis varies with environmental factors.

Districtwide socioeconomic status (or SES) is one feature of the school environment that is clearly related to a curricular emphasis on tested and basic skills. (See Table 29.) Teachers working in low SES communities find more need to stress tested skills in their classrooms and more stress on basic skills in their schools than those working in higher SES districts. At the elementary level, this response trend is statistically significant. It appears, then, that testing is driving the curriculum in economically disadvantaged areas to a greater extent than elsewhere, particularly in elementary schools.

Table 29. Teachers' Perceptions of the Emphasis on Tested and Basic Skills, By District Socioeconomic Status (SES)*

District SES Ranking[1]	ELEMENTARY[2]	SECONDARY[3]
High	10.41	9.52
Middle	10.35	10.13
Low	11.46	10.36

**Explanation.* The values on this scale range from 4 (perceive *no increased emphasis* on tested and basic skills) to 16 (perceive *greatly increased emphasis* on tested and basic skills).

The scale shows the mean (or average) combined responses of teachers in category to the four statements listed in Table 27 under the heading, "Effects On the School" (pages 66 and 67). On each of the four statements, 1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree.

[1] The Orshansky Index was used as an indicator of school district socioeconomic status.

[2] Differences among groups are statistically significant at $p < .01$

[3] Differences among groups are not statistically significant.

If this is in fact occurring, what accounts for it? Is it simply the belief that students from low SES backgrounds need more learning time than others on the basic skills that tests cover? Perhaps, but other forces seem to be at work here, too. Principals in lower SES schools report paying more attention to test scores than those in higher SES schools. They count the results of standardized batteries, state assessment measures, and district-objectives-based tests as more important for informing district officials, the public, and parents about school achievement

(Table 11 page 35). In addition, districts more often establish specific test-score goals for lower SES schools. (Principals in 40% of these school report that their districts do so, while only 10% of the principals in higher SES schools do.) At the same time, however, national studies repeatedly show that students from lower SES background do less well on tests than peers who are more well-off. Thus, in lower SES schools, where more students have difficulty on achievement tests, achievement-test scores seem to count for more, to be more consequential. This can help to explain why, if the teachers responding are correct, educators in lower SES schools spend more time and resources than others on teaching the material that tests cover.

In states where minimum competency (proficiency) testing is required for promotion and/or graduation, high-school teachers note a significantly greater emphasis on tested and basic skills. To a greater extent than secondary teachers elsewhere, they find that more school resources are devoted to basic-skills subjects, that they must spend more teaching time preparing students for tests, and/or that they must focus instruction on the skills that minimum competency tests cover. (See Table 30.) (For some illustration of these phenomena, review the last set of interview comments, quoted on page 74.)

Table 30. Teachers' Perceptions of the Emphasis on Tested and Basic Skills, By State Minimum Competency Testing (MCT) Requirements*

STATE REQUIREMENT	ELEMENTARY[1]	SECONDARY[2]
MCT required for promotion/graduation, state-mandated measure	10.81	11.06
MCT required for promotion/graduation, local choice of measure	10.17	10.13[1]
MCT required for diagnosis, state-mandated measure	10.58	9.91
MCT required for diagnosis, local choice of measure	10.11	9.40
No MCT required	10.79	9.99

**Explanation*. The values on this scale range from 4 (perceive *no increased emphasis* on tested and basic skills) to 16 (perceive *greatly increased emphasis* on tested and basic skills).

This scale is the same as that in Table #29. See footnote to Table #29 for further explanation.

1. Differences among groups are not statistically significant
2. Differences among groups are statistically significant at $p < .01$

This same response pattern is not evident among elementary teachers. Those in states requiring minimum competency tests for promotion and/or graduation do *not* perceive a greater tested-and-basic skills thrust in their curricula than teachers operating under other conditions. This may be because the potential consequences of strong minimum competency requirements are deemed less serious for students in the lower grades (no promotion) than for those in high school (no graduation).

Together with the findings regarding SES discussed in the previous section, those described here support the hypothesis that where test results have greater consequences, testing influences the curriculum more.

CHAPTER 6

THE SCHOOL CONTEXT AND CLASSROOM TESTING PRACTICES

A central goal of CSE's Test Use in Schools Study was to provide a national portrait of assessment practices and attitudes toward student achievement testing in schools across the nation. The four previous chapters have done that, with illustrations and elaboration from fieldwork in a number of schools and school districts. A second goal of the study was to address the question, "What factors seem to influence the assessment practices that currently exist in our nation's schools?" A framework for examining this question was introduced in Chapter 1.

One way in which the study tested that framework was by examining relationships between testing practices and viewpoints and environmental features external to the school, e.g., state and local testing requirements, federal and state programs, the nature of the school community and its students. The results of those analyses which produced statistically significant results have already been reported. In review:

- Secondary students in states *without* minimum competency or proficiency testing time spend a significantly greater amount of time each year taking classroom achievement tests than students in other states. Secondary students where minimum competency testing is required for promotions and/or graduation spend the least amount of time on classroom achievement testing.
- Teachers perceive a significantly greater emphasis on tested and basic skills in: (a) elementary schools in lower socioeconomic areas, and (b) high schools in states that require minimum competency (proficiency) testing for promotion at certain grade levels and/or for high-school graduation.

A second way in which the study sought to discover influences on testing practices and beliefs was by exploring relationships between and among test use patterns, attitudes toward testing and various school contextual factors. The latter included leadership practices in monitoring and supporting testing, teacher training and staff development, the presence of resources that support classroom testing, the organization of curriculum and instruction, and the presence of resources that facilitate instructional differentiation in the classroom. It begins with an explanation of the variables used in the analyses and then goes on to describe the relationships uncovered, highlighting those factors which were found to be significantly related to testing practices.

This chapter reports the results of this exploration. The chapter concludes with a conceptual model that integrates all the relational analyses conducted, a model that helps to explain patterns of test use in the nation's elementary and high schools.

The Variables In the Analyses

The analyses investigating relationships between and among test use, attitudes toward (or beliefs and perceptions about) testing, and school contextual factors employed variables developed by aggregating related questionnaire items. These variables and their derivations are described below.

Test use variables. Information on teachers' use of tests was derived from the survey questions described in Chapter 3. Use of four types of tests or assessment strategies were examined:

(1) *Use of Formal Testing*, including: standardized, norm-referenced tests; district objectives-based tests; and minimum competency tests;
(2) *Use of Curriculum-Embedded Tests*, including: placement, chapter, and unit and other tests "that come with the curriculum materials I use";
(3) *Use of Teacher-Made Tests*;
(4) *Use of Teacher Observations and Professional Judgment*, including: "my own observations and students' classwork," previous teachers' comments and grades, and previous teaching experience.

Teachers who responded to the survey rated the importance of each of these types of assessment for four different classroom tasks: planning, initial grouping or placement, regrouping or changing placement, and report card grading. (See Chapter 4 for details.) Thus, to determine teachers' overall use of each of the four assessment types listed above, their ratings of the importance of that type were summed across all four tasks. If, for example, they rated teacher-made tests as "critical" (value = 4) for all four tasks, they received a "score" of 16 for use of teacher-made tests. Or again, if they rated curriculum-embedded tests as unimportant (= 1) for planning, somewhat important (= 2) for initial grouping of students, and important (= 3) for re-grouping and grading, they received a score of 9, adding the four ratings, for use of curriculum-embedded tests. In the associational analyses, these scores were averaged across groups of teachers.

Beliefs and perceptions variables. Information on teachers' perceptions and beliefs (or attitudes) about testing were derived from survey questions described in Chapter 5. Based on confirmatory factor analyses, these questions were aggregated to create three "attitude" variables:

(1) *General Attitude Toward the Quality of Tests*: This variable was constructed by summing teachers' responses to the statements listed in Table 27 under the headings, "Quality of Tests" and "Value, Usefulness of Testing." This provided an overall index of the extent to which teachers felt testing was, on the whole, a good thing or a bad thing.

(2) *Perceived Emphasis on Tested and Basic Skills.* This variable was constructed by summing teachers' responses to the statements listed in Table 27 under the heading, "Effects on the School."

(3) *Attitude Toward Minimum Competency Testing.* This variable was constructed by summing teachers' responses to the two statements listed in Table 27 under the heading "Fairness, Desirability of Minimum Competency Testing."

The procedures for summing responses in building these scales followed those described above in the discussion of the test use scales.

School leadership in linking test results with instruction. This variable was built by summing teachers' responses (not principals') to the three statements listed under "The School Administration . . ." in Table 20, Chapter 4. It represents the regularity with which school administrators meet with teachers to examine the curricular and instructional implications of test scores, check to see that teachers follow up on these implications in their teaching, consider students' test results in teacher evaluation, and/or establish specific test-score goals for teachers to meet. Below, all this is glossed by the label, "Curricular Accountability," since it reflects the extent to which schools make curricular decisions based on test results and hold teachers accountable for these decisions.

Information and training about testing. Data on this factor came from teachers' responses to the items displayed in Table 23, Chapter 4, which asked respondents to indicate the kinds of informational and instructional activities their districts and schools had provided in the area of assessment over the past two years. Exploratory analyses sought to identify patterns in teachers' answers that would indicate types of staff development emphases, e.g., training programs that focused on improving teachers' skills at classroom assessment, in interpreting the instructional implications of test scores, on preparing students for testing, etc. These analyses showed no such patterns, however. In the end, this variable was constructed simply by totaling the number of different informational or inservice activities in which teachers said they had participated. Thus, it may represent the amount of attention paid to assessment issues in a teacher's school as much as it represents the depth of instruction teachers have received in testing.

Resources that facilitate classroom testing. Data on these resources was gathered through the questionnaire items listed in Table 24 of Chapter 4. The variable reflects how many of the four resources shown there (test item banks, computerized scoring, assistance in correcting and grading tests, collegial help in constructing tests) teachers have available and how frequently they use those that they have.

Resources that facilitate instructional differentiation in the classroom. In a set of questionnaire items not previously discussed in this paper, teachers were asked to indicate which of the following five human and material resources were available to them: (1) an aide, paraprofessional, or volunteer to assist with small group instruction or individual work; (2) other teachers with whom to divide up students "for extra help"; (3) instructional machines (audiovisual, computer, etc.) for independent work; (4) alternative curriculum materials for independent work to meet special needs (e.g., self-paced kits, etc.); and (5) specialists outside the classroom to whom students can be sent for special work. In addition to noting which of these were available to them, teachers estimated how frequently they used those that were. Thus, this aggregate variable was built by summing the number of the five resources a teacher used infrequently (several times a year or less, scores as "1") and the number used frequently (monthly or more often, scored as "2").

Students' total testing taking time, in terms of the total number of minutes spent annually as reported by teachers, was also considered in the context of these variables. Student's time on testing, however, was related to none of them; it is discussed no further here.

Some Relationship Between Testing Practices, Attitudes Toward Testing, and School Contextual Factors.

Correlations were run in a first analysis step to explore relationships between the variables just described. Table 31 shows the statistically significant results. As noted above, the information-and-trainingabout-tests factor reflects how much information and training teachers received through staff development activities in the last two years. It seemed reasonable to assume that knowledge about testing and about how test results can be used in the classroom could facilitate teachers' use of tests and/or influence their attitudes toward testing. The correlative analyses support these hypotheses, particularly at the elementary-school level. More training is associated with greater use of formal tests for instructional decisionmaking and with more positive attitudes towards the quality and utility of formal tests. (See Table 31.) Amount and diversity of staff development, however, are not related to the use of curriculum-embedded or teacher-made tests — probably because the kinds of inservice training teachers report usually focus on more formal measures, (Chapter 4, Table 23).

Curricular accountability is also related to test use and attitudes toward formal tests. Survey results indicate that when principals show that they care about test scores — by reviewing them to identify curricular weaknesses, taking action to assure teachers are emphasizing skills that test scores show are needed, etc. — teachers rate tests as more important in their instructional planning and, simultaneously, feel that tests are more valuable and useful.

Survey findings indicate that resources to facilitate classroom testing are not widely available (Table 24, page 55). Nevertheless, the greater the number that are available, the greater the importance teachers accord to all kinds of assessment results, including their own observation-based judgments.

The use of test results for instructional planning and decisionmaking assumes that some action can be taken on the basis of student test scores — e.g., providing remediation or advanced work for individual or small groups of students. Instructional resources, such as aides, instructional machines, and alternative curriculum materials must be available to make such actions feasible; where there are no options, no decisions are necessary and likewise test scores indicating the need for alternative actions are superfluous. Survey findings support this logic: availability of instructional resources is related to the use of all kinds of tests at the elementary school level and to the use of formal and curriculum embedded tests at the secondary level.

Table 31. Relationships Between Contextual Factors and Testing Practices

	STAFF DEVELOPMENT				LEADERSHIP SUPPORT				INSTRUCTIONAL RESOURCES				TESTING RESOURCES			
	Elem.		Sec.		Elem.		Sec.		Elem.		Sec.		Elem.		Sec.	
	R	M	E	M	R	M	E	M	R	M	E	M	R	M	E	M
Attitude Toward Quality of Tests	.318	.206	.215	—*	.230	—	206	—	—	—	—	—	—	—	—	—
Use of Formal Testing	.350	.300	.198	.256	.219	.235	.163	.333	.171	.288	.207	.230	.229	.340	.126	.220
Use of Curriculum Embedded Tests	—	—	—	—	—	—	.156	.376	.254	.391	.215	.236	.232	.361	.286	.237
Use of Teacher Made Tests	—	—	—	—	—	—	—	—	.206	.430	—	—	241	.362	—	.176

*Statistically non-significant ($p < 2.05$) correlations have been indicated with a '—'.

A Conceptual Model for Teacher Test Use

The previous section presented the results of a series of exploratory analyses designed to identify possible relationships between school contextual factors, attitudes toward testing and test use. This section examines these relationships within the framework of a single conceptual model that would examine all the influence on testing embodied in the study, i.e., both those in the immediate school context and factors external to the school, capturing important policy implications of the study. It should be stressed that while this examination was conducted using the techniques of path analysis, the results should not be construed as anything more than suggestive. Because of the exploratory nature of the analyses no formal tests of the conceptual model or of alternative models were conducted. Only single relationships (paths) were tested for statistical significance. Thus, while the model presented shows significant relationships between the constructs, it shows only one set of relationships, not necessarily the most powerful statistically. The remainder of this section is organized by the results of the path analyses for elementary and secondary teachers.

Elementary Teacher Test Use

The conceptual model shown in Figures 3 and 4 incorporates the results for four different "outcomes": teachers' use of formal tests, curriculum embedded tests, teacher-made tests, and teacher observations/judgments. For each of these, we examined the relationships between amount of use and the above variables including: attitudes about quality of tests, perceived emphasis on tested basic skills, school leadership in linking tests results with instruction, information about tests, testing resources and instructional resources and school level socioeconomic status. It was hypothesized the school SES would act as an exogenous variable in this system of relationships. Further, it was thought that school leadership in linking test results with instruction would influence the amount of information and training received by the teachers. That is, participants who were viewed as emphasizing and supporting greater use of tests were also likely to provide and require more training on test use. Lastly, it was assumed that leadership and information would relate to attitudes about test quality and basic skills press.

The tenability of these hypotheses can be ascertained from the results presented in Figures 3 and 4, displaying results of elementary school reading and mathematics. The paths drawn in these figures represent statistically significant regressions between the variables involved. Paths not drawn in the diagram indicate that the regression was not statistically significant.* Looking at the results in these two figures, one is struck by the high degree of correspondence. In fact, there is only one relationship that was statistically significant in one case and not the other. For elementary teachers there is a significant relationship between the amount of instructional resources and use of formal tests in math while that relationship does not appear for reading. With that exception the two models are identical in their structure indicating that the same mechanism is likely to be operating regardless of subject matter.

Beyond the concordance between the two cases there are several interesting features of the model. First of all, the influence of SES on the use of tests in decisionmaking is moderated through variables which are directly under administrative control. Specifically, the amount of information and training about tests and the degree to which the principal exercises leadership and holds teachers accountable, moderate the influence of SES on test use. Thus, regardless of a school's SES it appears possible through administrative steps to influence a teacher's use of tests. This administrative effect appears to be manifested through the attitudes that teachers have about tests. In particular, teachers seem to have better attitudes about the quality of tests in schools where there is more information and training about tests. Additionally, teachers who are more informed about tests and are held more accountable by the principal for test results also perceive a greater emphasis on basic skills and basic skills tests. These characteristics translate into greater use of formal testing in making classroom decisions.

The use of formal tests is also a function of the amount of resources available to the teacher. The greater amount of testing resources (e.g., scanning, scoring help) the greater the use of formal testing. Further, increased instructional resources leads to greater use of formal testing. The hypothesis here is that resources permit instructional alternatives or options. The existence of these options requires greater decisionmaking on the part of teachers and hence greater use of test results.

The use of curriculum embedded tests seems to be a function of the amount of both testing and instructional resources as well as the teacher's perception of the quality of tests. In situations where the teacher feels that the commercial tests are well made, they will be more likely be employed in decision-making. Again, the role of resources seems to be one of making testing or test use more feasible.

---

*A probability level of .05 was used in these analyses to determine statistical significance. The single exception to this criteria has been noted in the Figures. The basis for this exception was the exploratory nature of the analysis which generally involves somewhat more lenient criterial for examination of results.

It is interesting to see in the results of these analyses that the only contributing factors to the use of teacher-made tests and teacher judgment are the resources available to the teacher. This finding may reflect the pervasive use by teachers of these mechanisms for arriving at instructional decisions almost independent of other sources of information. That is, there may be a feeling on the part of teachers that their own tests and judgments are more suitable for decisions than more formal measures regardless of their attitudes and training about these latter tests.

In sum, the model portrayed in Figures 3 and 4 shows that the use of test information in teacher decisionmaking can be influenced by administrative action. In particular, the administrator can require greater accountability on the part of the teachers, provide more information and training about tests and, if feasible, supply additional testing and/or instructional resources. Each of these actions appears to positively influence the use of one or more types of test use.

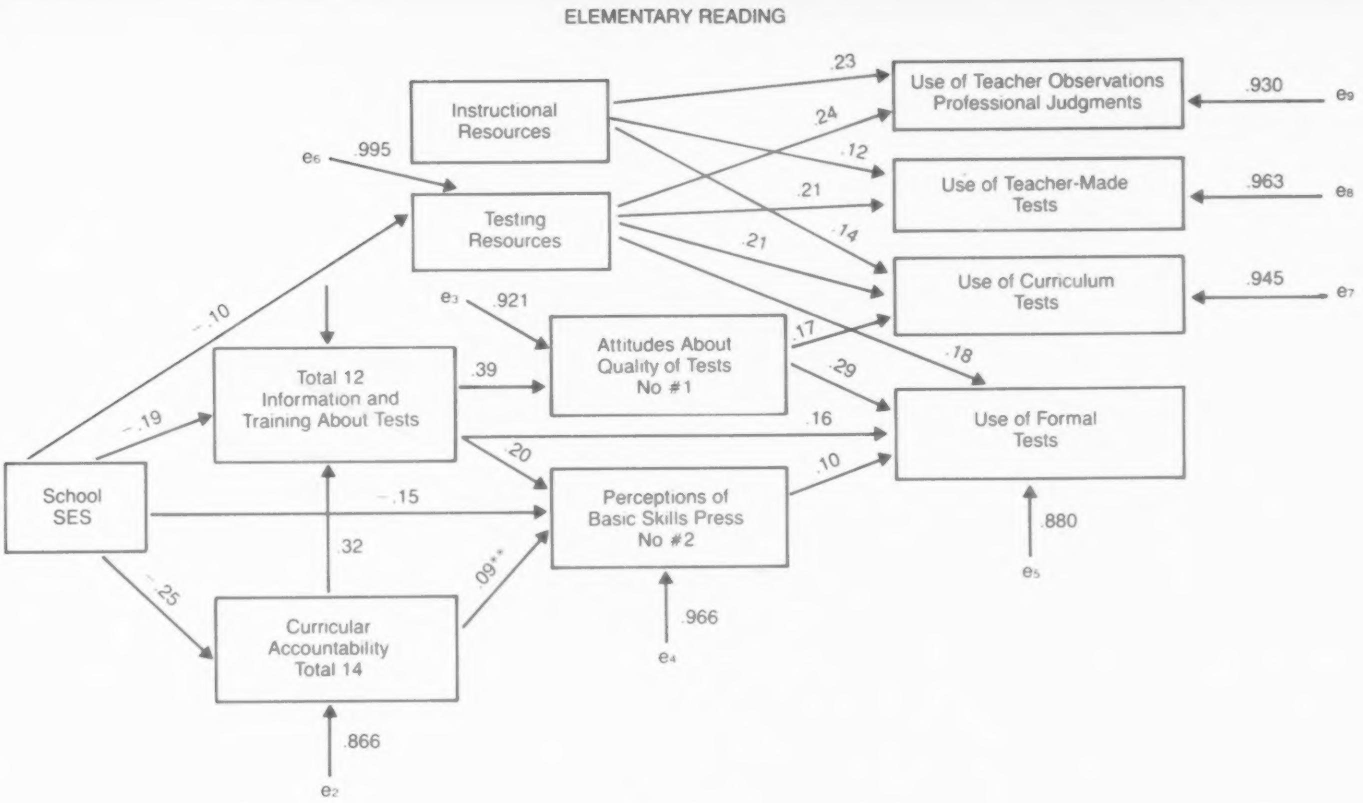

ELEMENTARY READING
Instructional Resources
Testing Resources
Total 12 Information and Training About Tests
School SES
Curricular Accountability Total 14
Attitudes About Quality of Tests No #1
Perceptions of Basic Skills Press No #2
Use of Teacher Observations Professional Judgments
Use of Teacher-Made Tests
Use of Curriculum Tests
Use of Formal Tests
e6
.995
e3
.921
e4
.966
e2
.866
e5
.880
e9
.930
e8
.963
e7
.945
−.10
−.19
−.15
−.25
.32
.09**
.39
.20
.16
.17
.29
.10
.23
.24
.12
.21
.14
.21
.18

Figure 3

Conceptual Model for Elementary School Teachers' Test Use in Reading*

*Reported values correspond to standardized path coefficients that were statistically significant (p < .05)
**Reported coefficient statistically significant (p < .06)

ELEMENTARY MATHEMATICS

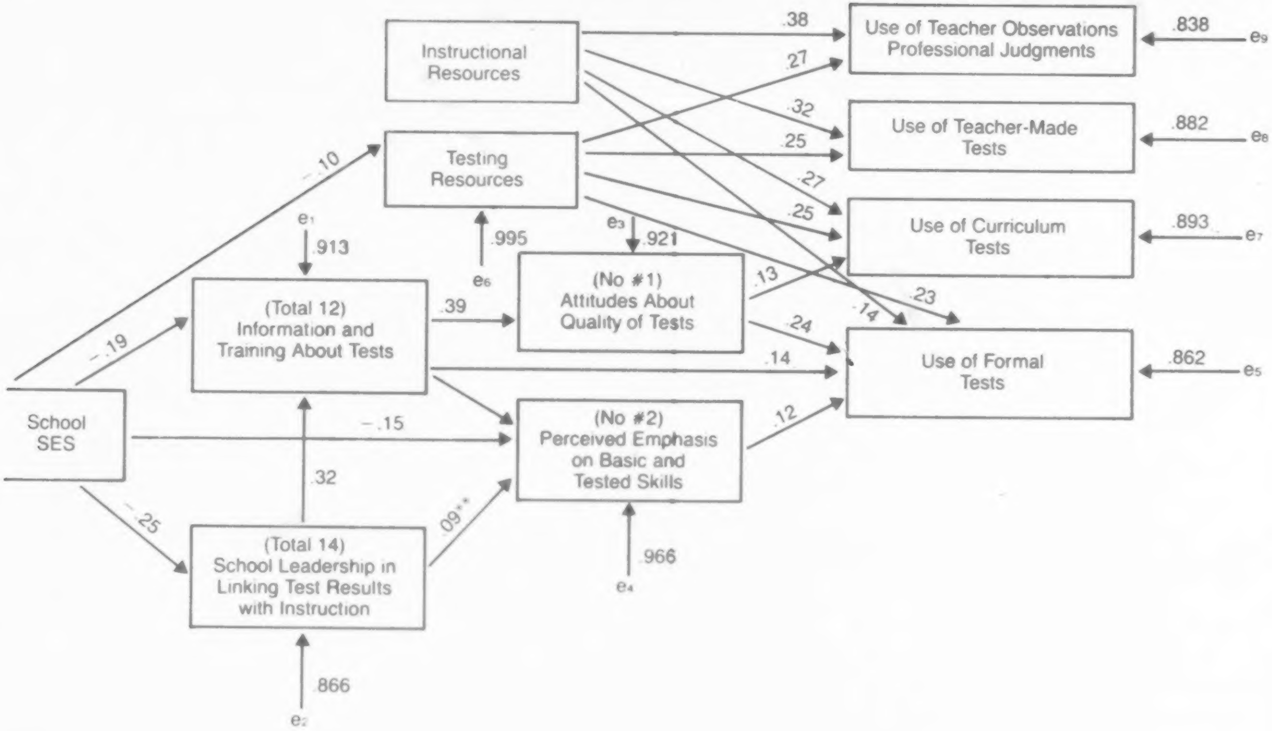

Instructional Resources
Testing Resources
(Total 12) Information and Training About Tests
(No #1) Attitudes About Quality of Tests
(No #2) Perceived Emphasis on Basic and Tested Skills
School SES
(Total 14) School Leadership in Linking Test Results with Instruction
Use of Teacher Observations Professional Judgments
Use of Teacher-Made Tests
Use of Curriculum Tests
Use of Formal Tests
.38
.27
.32
.25
.27
.25
.13
.23
.14
.24
.14
.12
.838
e9
.882
e8
.893
e7
.862
e5
−.10
−.19
−.15
−.25
.39
.32
.09**
e1
.913
.995
e6
e3
.921
.966
e4
.866
e2

Figure 4

Conceptual Model for Elementary School Teachers' Test Use in Mathematics*

*Reported values correspond to standardized path coefficients that were statistically significant ($p < .05$)

**Reported coefficient statistically significant ($p < .06$)

Secondary Teacher Test Use

Similar analyses were performed for secondary school teachers who taught English and mathematics. The results of these analyses are presented in Figures 5 and 6. As can be seen from these figures the picture at the secondary level is not nearly as clear nor consistent. In fact, there are few statistically significant relationships for the English teachers and those that do exist are for the use of curriculum tests. Because of the paucity of relationships for these teachers it would be hazardous to attempt to interpret them or the model.

The results for mathematics teachers are somewhat more encouraging though still not as conceptually appealing as the elementary school results. The results in Figure 5 show that a somewhat similar mechanism to that found in elementary schools may be operating for the use of formal and curriculum tests. That is, it appears that administrative leadership, information about tests, and testing resources are all influencing the use of formal and curricular tests. What appears to be different at this level, however, is the greater direct role of school leadership in linking test results with instruction. This variable has strong direct relationships to both use variables. Further, this variable, rather than information about tests, seems to relate to teachers' attitudes about test quality. Thus, these results seem to point to a greater direct role for the principal at the secondary school than at the lower grade levels. It should be noted, however, that the same constellation of factors are evolved, it is just their relative priorities and interrelationships that are different. Therefore, from a prescriptive point of view, working on the three variables of information and training about tests, school leadership, and testing resources seem most likely to pay off in terms of greater teacher use of formal and commercial tests.

In summary, these analyses have explored a possible prescriptive model for teachers' use of different types of information in their decisionmaking. While the results showed some disparity between elementary and secondary teachers, particularly for secondary English teachers, some definite similarities were found. In particular, it appears that three policy relevant and administratively manipulatable variables are related to increased use of formal and commercial tests. These three variables are the amount of curricular accountability operating in the school, the amount of information and training given to the teachers about tests, and the amount of testing-related resources made available to the teacher. It would appear that if increased use of formal test results were considered a desirable goal, increased emphasis should be placed in the three areas mentioned above.

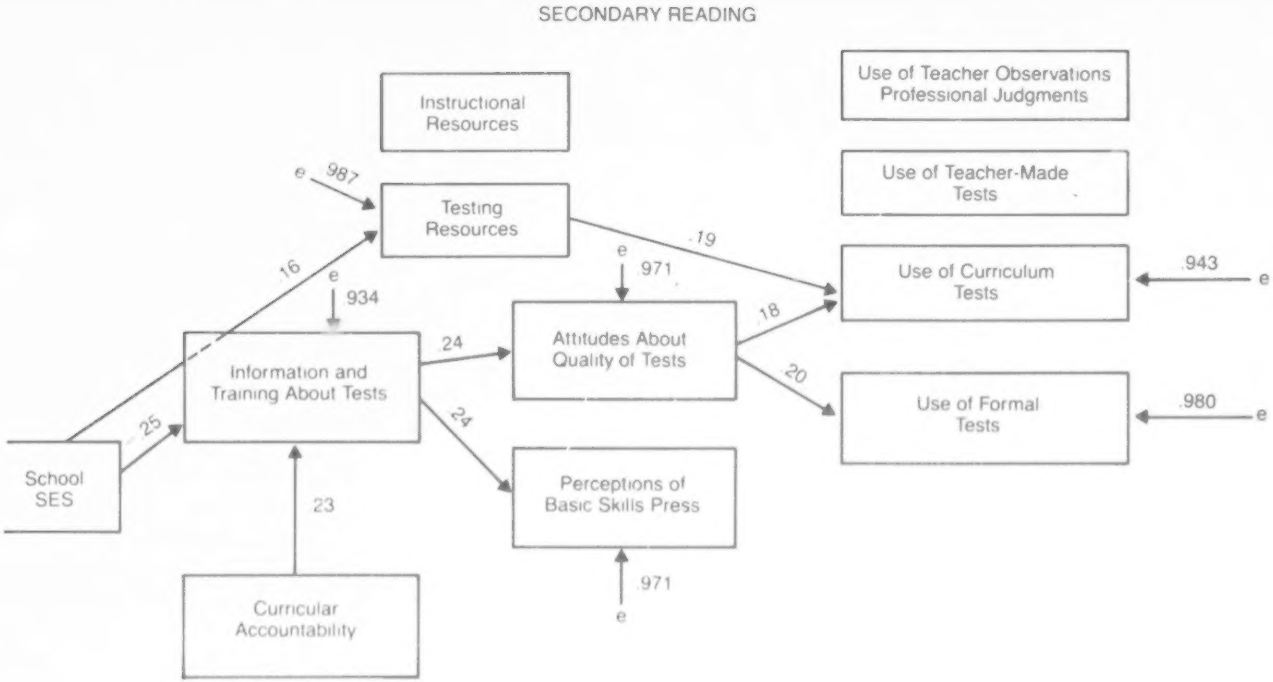
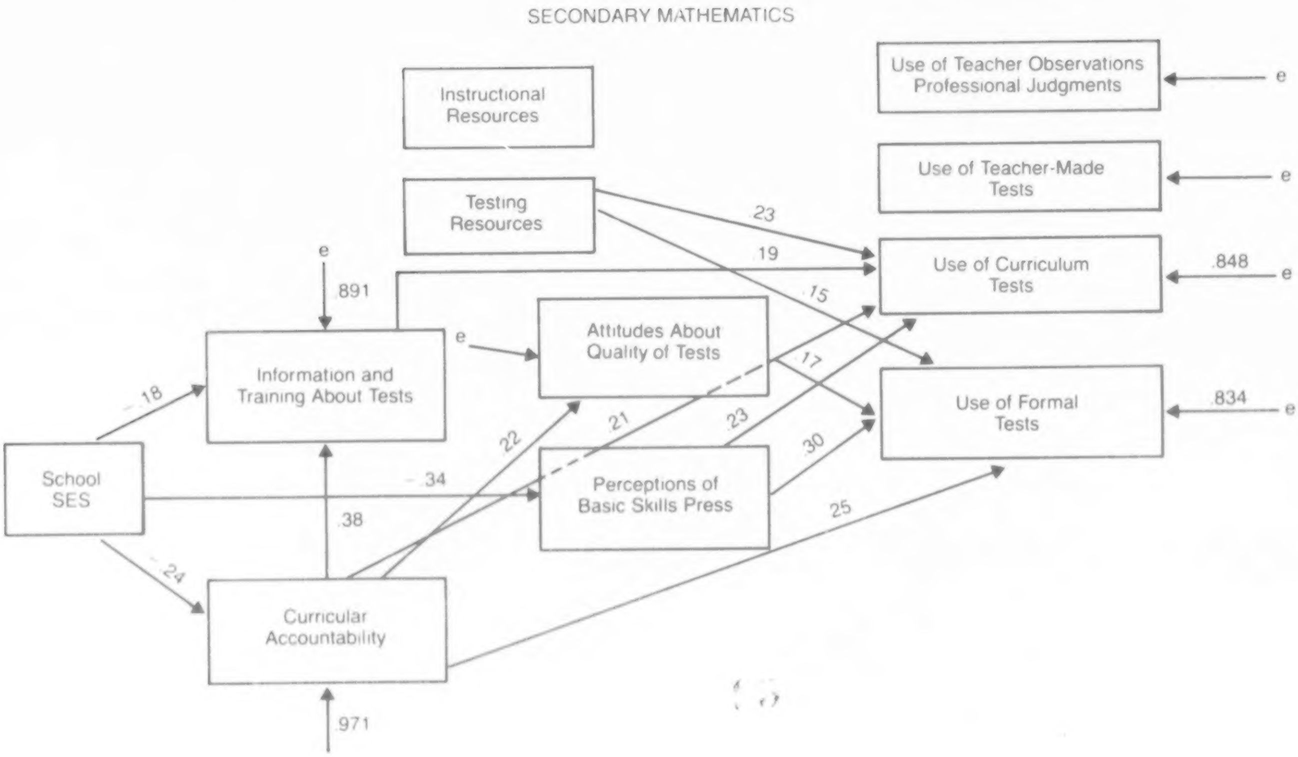

SECONDARY READING
Instructional Resources
Use of Teacher Observations Professional Judgments
Use of Teacher-Made Tests
e
.987
Testing Resources
.19
e
.971
Use of Curriculum Tests
.943
e
.16
e
.934
.18
.24
Attitudes About Quality of Tests
.20
Information and Training About Tests
Use of Formal Tests
.980
e
.25
.24
School SES
Perceptions of Basic Skills Press
.23
.971
e
Curricular Accountability

Figure 5

Conceptual Model for Secondary School English Teachers' Test Use*

*Reported values correspond to standardized path coefficients that were statistically significant ($p < .05$)


SECONDARY MATHEMATICS

Instructional Resources
Testing Resources
Use of Teacher Observations Professional Judgments
Use of Teacher-Made Tests
Use of Curriculum Tests
Use of Formal Tests
Information and Training About Tests
Attitudes About Quality of Tests
Perceptions of Basic Skills Press
School SES
Curricular Accountability
e
e
e
e
e
e
e
.23
.19
.15
.17
.891
−.18
.22
.21
.23
.30
−.34
.38
−.24
.25
.848
.834
.971

Figure 6

Conceptual Model for Secondary School Mathematics Teachers' Test Use*

*Reported values correspond to standardized path coefficients that were statistically significant ($p < .05$)

CHAPTER 7

SUMMARY AND IMPLICATIONS: ISSUES FOR STATE AND NATIONAL POLICY MAKERS

The findings of CSE's Test Use in Schools Study map the topography of basic-skills achievement testing and achievement test use in public schools across the United States. They show patterns of local assessment practice, demarcate the domain and scale of local leadership in assessment, and shade in the tones of local educators' beliefs about testing and its influences on their schools. Through its associational analyses, the study also draws some tentative lines between regions on this map. That is, it models some ways in which these within-school phenomena appear to be tied functionally to one another and to certain conditions beyond the schools.

This map was constructed, as Chapter 1 explained, with certain policy concerns in mind. Thus, it not only describes the landscape of public school achievement testing; it also illuminates it such that: (1) some issues and concerns particularly important to national and (especially) state policy makers stand out in relief; and (2) some answers to local policy makers' questions become clearer.

After an interpretive review of study findings that frames the discussion of both these sets of policy issues, this chapter outlines three findings that fall in the first category listed above — those most appropriately addressed at the state and national levels. One is the matter of equity in testing, as raised by study findings regarding the impact of required tests. The second is the issue of teacher preparation and local test quality, as raised by findings of this and related studies. The third is the critical need to explore ways of integrating, aligning, or rationalizing assessment such that the same or similar test data can be aggregated to address the diverse needs and multiple questions of policy makers at various hierarchical levels in the nation's educational system, e.g., in the classroom, the school, the district, the state, and the federal government.

In the next chapter, case study data elaborate survey results and suggest concrete answers to questions of test utilizations and testing efficiency at the local level. More specifically, that chapter demonstrates some ways in which district administrators can act to achieve collective links between testing and instructional decisionmaking.

Summary: The Study Reveals Two Tiers Of Achievement Testing, Both Under-Utilized.

A close examination of Test Use in Schools Study results confirms that there are two tiers or layers of student-achievement assessment in our schools today. These are consistently distinguishable from one another in their proprietorship, characteristics, and functions. One tier of assessment is internal or local to the schools. It is "owned," and for the most part produced, by teachers themselves. This local or internal tier includes two main types of assessment: (1) the tests, quizzes, and other measures that teachers construct and administer in the course of their teaching, and (2) the clinical judgments of students' achievement that teachers form as they interact with students and observe their work in various classroom situations day after day. A third kind of measure also figures in this tier, but it is especially important for elementary-school teachers. These are the tests included with commercial curriculum materials used in the classroom. While these are not produced in the school, teachers in the elementary grades are most often invested in them. Teachers often have a say in choosing (and choosing how much to use) them and the materials they accompany; teachers can time their administration and adapt their content to fit the pace and emphases of instruction.

The second tier of assessment is external to the school: mandated by the district, state, and/or suggested by federal program requirements (e.g., for placement in compensatory education programs). Norm-referenced, standardized test batteries are the most common among these. Other types of measures used for minimum competency (or functional literacy) testing or as part of state assessment programs are also included here. In some cases, too, tests constructed or purchased by districts and referenced to their curricular objectives fall in this second category. Tests of these kinds are developed beyond the schools. Their administration is called for primarily to meet organizational needs and concerns at higher levels of public-education governance. Those who work at those levels may have a sense of ownership in these tests; educators in the schools rarely do.

These two tiers of assessment function quite differently in most schools and districts. Teachers and principals rely heavily on the results of internal assessment strategies and consider them important as they go about routine instructional planning and decision making. At the same time, they generally treat information from external testing as of minor importance using it only occasionally and idiosyncratically. These patterns are obvious in both CSE's fieldwork findings and survey data.

When teachers were interviewed during pre-survey fieldwork, they discussed all the information they had throughout the year on students' academic capabilities, performance, and progress; they described whether and how they used that information. Collectively, they cited far more uses for the information that came from assessment strategies that were local to the school and classroom. (See Table 17, page 42.)

Teachers surveyed across the nation were asked to rate the importance of diverse types of assessment results in four routine, decisionmaking tasks. Again, the pre-eminence of the internal tier of assessment was apparent. (See Tables 12 and 13, pages 36 and 37.) Principals in CSE's national survey were asked to rate how important a role data from various sources played in eight regular school-level administrative activities. Here, the separate functions of the two tiers of achievement assessment were especially apparent. Principals reported counting internal assessment data more heavily in making instructionally relevant decisions, e.g., allocating funds, assigning students, evaluating teachers. But they indicated that results of external measures were more important in reporting to those beyond the school, e.g., to district administrators and the public. (Review Table 10, page 32.) Further evidence of the functional independence of the two tiers of student-achievement assessment appears in Figures 3 through 6 of Chapter 6. In general, these figures show two networks of relationships. One includes the use of measures external to the school (formal tests); the other, internal assessment techniques (teacher made tests, teacher observations and professional judgments). The use of tests in the external tier varies in response to a chain of factors that usually includes the perceived need to emphasize tested and basic skills ("basic skills press"); administrators' holding their teachers accountable for test-score-based curricular decisions ("curricular accountability") attitudes about test quality; and information and training about tests. None of these factors, however, influence the use of the two most widespread types of school-based, or internal, assessment — teachers' tests, observations and judgments. Instead, teachers' use of the latter is tied only to classroom circumstances: to instructional resources that permit differentiated instruction to meet students' individual learning needs and (less strongly) to resources that save time in testing. (The single exception is in high-school English classrooms, Figure 4, where teachers' use of local measures does not covary with any of the factors included.)

These findings suggest that external test results become more important to teachers only when something or someone impels or induces teachers to treat them as more important. Instructional circumstances do not influence teachers' use of these results. On the other hand, the results of internal assessment techniques *are* influenced by instructional assessment circumstances. When classroom conditions demand and facilitate closer, more fine-grained evaluation of students' performance, it is their own, local measures that they weigh more heavily.*

Taken together, the research findings just cited show that there are notable quantitative differences in the ways the external and internal tiers of assessment are used by educators in the schools. They reveal that the results of externally mandated testing serve fewer purposes (Table 17) and are not counted as heavily in planning or decisionmaking (Tables 9 through 13). But fieldwork clearly suggests that there are also significant qualitative differences in how the two tiers of assessment are typically utilized by teachers and principals. The results of external tests are most often examined briefly, casually, and asystematically. Do principals consider the results of standardized and district-objectives-based tests in curriculum evaluation? Table 9 suggests that they do. But interviews indicate that this often means that they merely glance over the scores, mention them in a faculty meeting, and point out the areas in which the school did especially well or poorly. (See quotations, page 84 in Chapter 5.) Do teachers use standardized test results in planning? Apparently they do to some extent (Tables 1 and 2). Fieldwork suggests, however, that, more often than not, this means a once-a-year visit to the office for a quick look at their students' cumulative files. Are standardized test batteries and minimum competency scores consulted in student placement? Again Tables 9, 12, and 13 indicate that they are. But visits to schools make clear that they are most often consulted as part of an automatic or cursory gate-keeping procedure. Law or policy guidelines direct that students with scores below a certain cut-off point be placed in a compensatory program or remedial class. Alternatively, as one high-school teacher put it, describing a procedure reported by many:

---

*Note that the use of curriculum-embedded tests, considered here as internal measures, tends to fall between or overlap the two relational networks described above. Nevertheless, use of these tests generally correlates more strongly with classroom instructional and testing resources than with the factors that influence external tests.

> They give me each kid's standardized-test score on my class roster. If one stands out, I usually check with the counselor to be sure the kid should really be assigned to geometry.

Such uses contrast sharply with teachers' recurrent and systematic use of assessment techniques that are local to the classroom and school in an on-going process of intructional planning and decisionmaking. They contrast markedly with principals' serious consideration of teachers' advice, recommendations, and grades on teachers assignments in making budgetary decisions or next year's class assignments. And they certainly do not constitute thorough utilization of external testing data in a systematic process of school-wide analysis decisionmaking, or planning of curriculum and instruction.

Why do the two tiers of achievement assessment function in the different ways that they commonly do? The reasons are not hard to find. They lie in the interplay of several factors: characteristics of the measures themselves, circumstances surrounding their availability, educators' training in assessment, and the organization of educational planning in schools, districts, and beyond.

American educational organizations (schools, school districts, etc.) have been called "loosely coupled systems" (c.f., Deal, 1979; Meyer & Rowan, 1978; Montjoy & O'Toole, 1979). Schooling in the United States has been described as "pre-industrial — a cottage industry" (Dawson, 1977). And teachers in classrooms have been likened to "street-level bureaucrats" (e.g., Weatherly & Lipsky, 1977). These similes call attention to the relative autonomy of the classroom teacher in multileveled decisionmaking hierarchy — a hierarchy in which participants at each level have interests and concerns that only partially overlap, only sometimes coincide.

For their part, teachers routinely do a great deal of instructional planning. They have a major role in planning what to teach (and/or emphasize) and how to teach it, in diagnosing individual students' learning needs, and in assuring that students are working at appropriate levels in the curriculum. As the school year unfolds, they need to monitor their students' progress, to consider whether and how to adjust the pace and emphases of their teaching, to grade students and inform parents of achievement-to-date, and so on. To do all this and do it well, teachers need assessment tools wth three basic characteristics: (1) *Validity* — they must assess what the teacher believes he or she has actually taught in a way that seems consonant with the way he or she has taught it; (2) *Suitability* — their intended purposes must fit the tasks the teacher needs to accomplish (thus teachers seek placement tests for placement, chapter and unit tests for monitoring progress and grading, etc.); and (3) *Immediate Availability* — the teacher must be able to employ them whenever it seems appropriate to do so and have the results back promptly. In short, the assessment tools that teachers need must be sensitive to local conditions, to the array of particular circumstances in their particular classrooms at the moment. And, in order to function throughout the year as the instructional leaders of their schools, principals need measures of the same kind.

It is not surprising, then, that both teachers and principals rely heavily on assessment strategies that are internal to the school and its classrooms; teacher-made tests and assignments, teachers' observations and clinical judgments, and the adaptable, readily available tests that come with the commercial curriculum materials they are using. From their points of view, these internal measures have all three of the characteristics listed above. Externally mandated measures, on the other hand, usually do not. They are not designed primarily to provide data for routine classroom decision making. The fit between their contents and format and a particular teacher's curriculum is problematic. Often, their scores are not returned until weeks or months after administration. Often too, the results come back in a format teachers and many principals find unfamiliar and/or cumbersome. (See Table 19, page 48.) For any or all these reasons, the results of standardized tests, other minimum-competency measures, and many district-objectives-based tests can seem remote and irrelevant to teachers and principals. In addition, teachers and principals generally have limited formal training in testing and measurement or the use of test data. (See Table 23, page 54.) Further evidence that supports this claim will be found further on in the chapter. This also limits the accessibility of external testing data to educators in the schools. CSE's Test Use in Schools Study fieldwork found teacher and principals voicing these very concerns as drawbacks of external testing. (See illustrative quotations in Chapter 5, pages 68 and 69).

But the very characteristics that make internal assessment tools ideal for use in individual teachers' and principals' routine work severely restrict their utility for systematic school- and district-wide planning. Their content and the timing of their administration is idiosyncratic, variable from classroom to classroom. Aggregating the data they provide in order to see achievement patterns across grade levels, a department or the entire school, therefore, is difficult if not inappropriate and impossible. This is especially true of teacher-made tests and assignments, but it also often applies to tests embedded in texts and other commercial materials. (Teachers time their administration differently; they sometimes adapt their contents. The same materials or text series are not always used throughout the school.) And while teachers' cumulative observations and experience-based judgments are valuable sources of information, they cannot be readily synthesized into a precise, detailed, picture of specific curricular or teaching strengths and weaknesses across many classrooms or schools.

It is these problems with local or internal assessment strategies that have made standardized, minimum-competency, and special district-objectives-based tests attractive to local school districts — and make similar measures a virtual necessity for states and other educational agencies. By providing standard and consistent data across settings, such tests facilitate comparisons among classrooms, schools, and/or districts; they permit year-to-year monitoring of performance. They are likely to be more sound psychometrically than teachers own tests; in most circumstances they are sufficiently valid to indicate broad patterns and trends. Tests of these kinds can take time to administer, score and analyze comprehensively, but comprehensiveness is important to district and state planning, especially if data are gathered only annually or biannually. Coming full circle, however, the same features that make these types of measures useful to districts and larger education agencies generally limit their usefullness for teachers and principals. Thus, two tiers of achievement testing, largely distinct in their functions, are maintained in public schooling.

As noted earlier, the next chapter will present research-based models and guidelines detailing how districts and schools can begin to integrate these two tiers of testing use both more fully in planning for instructional improvement. The remainder of this chapter, however, goes on to examine three important issues that their separation raises of state and national policy makers.

External Assessment: Study Findings Raise Issues of Equity

Chapter 1 explained some of the mechanics through which formal, mandated tests (the external tier of assessment) can serve as interventions, or agents of educational change. (See pages 4 and 5.) With this "testing as an intervention" hypothesis in mind, CSE sought to identify whether tests required by agencies beyond the school are in fact influencing school programs and so students' educational experiences and life chances. Among the policy questions underlying the Test Use in Schools survey (Chapter 1, pages 2 and 3) several addressed the influence of minimum competency testing: What are the impacts of different kinds of minimum competency programs? Have they affected curriculum and instruction? Have they wrought changes in the other ways that districts and schools measure student achievement? A second set of policy issues were raised about the formal testing (most often standardized, norm-referenced testing) occasioned by the evaluation requirements of state and federal education programs: How does such testing affect the instructional time of participating students? How does it influence the distribution of instructional staff members' energies and efforts?

Answers to these questions have been offered through the preceding chapters. Here, it is appropriate to review them and to extrapolate their implications.

Minimum competency testing: three potential sources of education inequity. Study findings raise the possibility that differential minimum competency or proficiency requirements from state to state (and in some states, from district to district) are generating educational inequities.

First, there is reason to question whether the tests in use are uniformly fair. Substantial percentages of teachers, especially in the elementary grades and high-school English, think that they are not (Table 25, page 60). Furthermore, where laws now specify competency tests as prerequisites for promotion to certain grades and for high-school graduation, both elementary and high school teachers are signficantly more inclined to doubt their fairness and the wisdom of using them as gatekeeping measures. (Table 28, page 73.) Put another way, those teachers in the best position to know the tests and to judge how well they function in sorting minimally competent from incompetent students are the very teachers most likely to doubt their equity and desirability.

Most teachers, of course, are not experts in testing and measurement. (See the discussion below, pages 103 to 106.) Their judgments of test fairness cannot be taken as definite. Nevertheless, the patterns of their survey responses should be sufficient to stimulate policy makers' continued concern about such issues as the instructional validity and cultural linguistic bias of the proficiency or minimum competency test now in use.

Second, survey results indicate that competency or proficiency testing may be generating differences in the frequency of routine classroom assessment in high schools. This, in turn, may be producing inequities in the quality of instruction. In *secondary schools* where no state-mandated competency tests exist, students spend roughly 62 hours a year taking English tests and 53 hours a year taking mathematics tests (Table 6, page 20). Given that tests in these subjects average about a half-hour each (Table 3, page 16), this means that the typical student in these schools takes an English test on the average of three times a week and a math test on the average of two-to-three times a week through 37 weeks of instruction each year. Where proficiency or competency tests are required for promotion and/or graduation, however, high-school students average half-hour tests in each of these subjects once a week or less across the school year.*

No one knows what the optimal of testing is, and some would argue that testing should be minimized to "save" class time for teaching and learning. A number of studies, however, indicate that frequent monitoring of student progress is an important characteristic of more effective schools. (See Purkey & Smith, 1982, for a comprehensive, critical review.) Combined with CSE's survey findings, this suggests that policy makers in both states and districts should be concerned about the direct and indirect effects of minimum competency requirements on local assessment practices. Whether and how these requirements influence classroom testing should be closely examined; research should explore how often testing should optimally occur. But if frequent monitoring of students' progress and prompt feedback on student performance are features of effective teaching, differential competency mandates may be contributing to inequities in the quality of students' instruction from one state to another.

Finally and perhaps most importantly, survey results raise the possibility that minimum competency or proficiency testing programs are working to produce state-to-state differences in the breadth of the curriculum that students experience, especially at the secondary level. There is substantial evidence that examinations with important consequences tend to influence the curriculum in schools where they are given (e.g., Cronbach, 1963; Linn, 1983a, b; Madaus & McDonough, 1979; Tinkleman, 1966). It is hardly surprising, then, that teachers in high schools where minimum competency tests are required for graduation agree, to a significantly greater extent than teachers elsewhere, that these tests affect the amount of time that they can spend teaching subjects and skills the tests do not cover, that they have recently been spending more teaching time preparing students for required tests, and that the proportion of their schools' resources allocated to basic skills teaching is so great as to detract from the quality of their overall educational programs. (Refer to Table 30, page 76 in Chapter 5.)

---

*In states that require tests for promotion/graduation and mandate the measure that schools must use, the averages are a classroom English test 1.2 times per week and a mathematics test once in every seven school days. In states that require tests for promotion/graduation but permit districts to select or design their own measures, the average is a classroom test every seven to seven-and-a-half school days in both English and mathematics at the secondary level.

Some maintain that tests should influence the curriculum. Linn (1983a, p. 125), for example, takes the position that

> a test provides the means of making agreed-upon objectives clear and precise. An important goal of instruction should be the achievement of those objectives as demonstrated by performance on the test.

Especially in the case of minimum competency in the basic skills, there are many who would agree. Educational policy makers and practicing educators, they would argue, should establish clearly and precisely the basic proficiencies they expect students to have at various milestones in their schooling. Instruction should work toward the achievement of these minimal objectives, and students should demonstrate that they have attained them through test performance. Indeed, it was arguments such as this that promoted the passage of minimum competency, proficiency, or functional literacy testing legislation in over 40 states.

Few would quarrel with the idea that students should attain minimal standards of proficiency in basic skills. But if the perceptions of teachers surveyed by CSE are accurate, requiring minimum competency testing for promotion and graduation may be narrowing the secondary curriculum: inducing districts, high schools, and individual teachers to emphasize the tested, basic, functional literacy skills at the expense of other learning. Thus, those students in states with such competency requirements may be limited to learning less about advanced composition, and less of the analytic and problem-solving skills that these subjects entail than students in other states with different requirements — and less than they themselves might be learning were their teachers not spending class time working to assure that everyone is proficient in the minimum, tested skills. Perhaps, then, these students — many of whom would certainly pass minimum competency tests in any case — are being placed at a disadvantage as compared to students in states where proficiency testing is not required or required only for diagnostic purposes.

Of course secondary students who fail proficiency tests where there are graduation requirements are more likely than others to experience a contracted curriculum. Fieldwork indicates that they are often placed in special remedial courses centered on the skills that the tests cover. The creation of such courses, however, can mean that fewer sections of more advanced courses are available for other students. (States have not always provided additional findings for remediation to accompany competency legislation; districts cannot always hire the extra teachers that would be needed to both maintain current course offering and staff remedial sections.) And while it is certainly important to make sure failing students gain minimal competence in basic reading, writing, and mathematics skills, it is also important to recognize that these skills in themselves do not open many doors in an increasingly high-technology society.

In short, CSE's survey findings raise serious questions for policy makers about the cost-benefit trade-offs of competency testing requirements, as well as questions about their equity. The tests may be unfair for many students. They may be reducing the frequency of routine classroom testing and (thus) the quality of instruction. They may be narrowing the curriculum and, with it, the range of opportunites open to many students. These possibilities deserve the attention and investigation of all those who shape educational policy at the local, state, and national levels.

Testing for state and federal program requirements: additional equity issues. Study findings also suggest that testing conducted to meet the evaluation requirements of federal and state educational programs may be influencing the educational experiences of low-income students at the elementary level.

According to principals' reports, the results of formal tests carry more weight and have greater consequences in schools serving low socioeconomic status (SES) neighborhoods than in those serving higher SES communities. In the former, they count far more in such tasks as planning curriculum, deciding on students' class assignments, allocating school funds, and reporting to the public, district officials and parents. (Refer to Table 11, page 35.) The role played by formal tests in these low-income schools is often mandated or enhanced by the special state and federal education programs in which they participate. Standardized, norm-referenced scores are commonly used in low-income schools, for instance, to establish individual students' qualifications for compensatory education programs. Formal testing plays a part, too, in the placement of non-English-speaking and limited-English-speaking students (many of whom came from lower SES families) in bilingual programs. These and similar programs usually entail evaluation requirements, and these requirements are frequently met through formal testing. Thus, as noted earlier (Chapter 3), federal and state program requirements help to make test scores especially salient in the very schools where more students more often have difficulty doing well on formal tests. And, to a significantly greater extent than others, teachers in lower SES schools find a greater need to spend classroom time on tested, basic skills and preparing students for required tests. They are also signficantly more inclined to agree that the measures allocated to basic skills instruction are so great as to affect the overall quality of their schools' programs. (See Table 29, page 75.)

Certainly all of the emphasis placed on test scores in low SES schools cannot be traced to the presence of state and federal program requirements. Nor can the greater attention given tested, basic skills in these schools be ascribed solely to their emphasis on test scores. Nevertheless, as noted in the last section, tests with important consequences can and do influence curriculum, and it is clear that state and federal program requirements do help to make test results more consequential in low SES neighborhood schools. Thus, those who establish the requirements for state and federal programs should give careful consideration to the role additional emphasis on test scores may play in narrowing the curricular opportunities of low-income elementary students, which can only add to the disadvantages such students already encounter.

Internal Assessment: Test Use in Schools Study Findings and Related Research Raise Issues of Teacher Preparation and Test Quality

While CSE study findings on the external tier of assessment (or formal testing) raise educational equity issues for policy makers, results regarding the internal tier of assessment generate concerns about test quality and teachers' training in assessment.

The formal tests mandated by agencies outside the school often play a role in major gatekeeping decisions regarding students. But teacher-made tests, teachers' daily assignments, and teachers' observations and judgments, play at least as great a role in influencing students' educational experiences and life chances. Constituting the tier of assessment internal to the schools, the results of these techniques are critical in schoolwide decisionmaking. They influence curricular planning, the distribution of school funds, and students' assignment to classroom. They also weigh heavily in what schools tell parents about their children's progress. (Review Tables 9 and 10, pages 31 and 32.) They are equally important in the classroom. They help to shape teachers' planning as the school year begins, significantly affect their placement of students in learning groups, and count most in their calculations of students' report-card grades (Tables 11, 12, 13, and 16 in Chapter 3). Thus, the various teacher-designed strategies of achievement assessment cumulatively shape students' learning environment, academic self-concept, educational status, and (ultimately) their socioeconomic opportunities.

Despite the obvious importance of teachers' tests, assignments, and clinical judgments, studies have repeatedly shown that teachers receive little pre-service training in assessment. Reviewing some of this literature in a recent paper, Coffman (1983) wrote:

> In 1959 Mayo reported a study by Noll indicating that 83% of 80 colleges he had surveyed offered a course in measurement, but that only 14% of them required one of all teacher education students. Furthermore, only 10% of the states required a course for certification. Ten years later Stinnet (1969) made no mention of any requirement in educational measurement in his encyclopedia article on teacher certification, nor did Burden (1982) thirteen years later. It seems obvious that only a minority of teachers have had any intensive training in educational measurement.

Recent research also indicates that teachers remain poorly prepared in assessment (Rudman, et al., 1980; Woellner, 1979; Yeh, et al., 1981). And as CSE's survey indicates, in-service training does little to fill the gap. Only about one-fifth of the teachers responding received staff development related to selection and construction of good tests or in use of test results to improve instruction.

Very little direct information is available about the quality of teacher-developed tests. As the previous paragraph should suggest, however, that which is available reveals that teachers lack skill in test construction. Ebel (1967) identified a variety of common errors in teachers' tests and urged better training in this area. In a recent review of teacher-made tests, Fleming and Chambers (1983) found that teachers write more questions of the short-answer kind than of any other type; they rarely devise essay examinations. For the most part, too, the tests reviewed required students to recall facts and terms. Questions requiring learners to translate, apply, or otherwise use knowledge were rare. Furthermore, Fleming and Chambers discovered a "general tendency" to omit test directions, to use illegible test copies, and "to omit the point values to be assigned to test questions. This trend suggests that teachers may not be visualizing their tests as means for quantifying students' performance as a measure of students' learning. This trend appears to confirm reports in the literature. . .that teachers' knowledge of fundamental measurements concepts is limited" (Fleming and Chambers, 1983, p. 36).

All in all, it seems worth considering just how qualified today's teachers are to be developers of the tests that most affect students' lives.

How effective are teacher-generated tests in revealing the insufficiency in individual students' learning? How valid are they as measures of students' achievement? How do teachers decide how often to test? How skilled are elementary school teachers in analyzing the commercial curriculm-embedded tests that they frequently use? Similar questions can also be raised about teachers' skills in making observation- and interaction-based judgments of children's learning.

Given the time spent on teacher-constructed tests and given the cumulative importance both of these tests and of teachers' judgments in classroom and schoolwide decisionmaking, teachers' preparation for the role of achievement assessor and their competency in that role needs thorough review. And this review deserves the attention of both the educational policy and the educational testing communities.

Toward More Integrated And Rational Assessment Systems

While they work to examine and (as necessary, rectify equity and quality problems in our current system of achievement assessment, policy makers will be well advised to explore ways for integrating that system and making it more national.

As the opening of this chapter explained, Test Use in Schools Study findings reveal national testing practices which are bifurcated by internal and external needs, replete with overlapping requirements at the federal, state and locals levels. The result is two systems or tiers of testing which are redundant and inefficient. Furthermore, survey findings show that significant teacher and student time is spent in required testing, representing fully half of the testing at the elementary school level and one-quarter of the total student testing time at the seondary level. This time presumably serves the decisionmaking and accountability needs of policymakers, but (as study results clearly show) serves very little the information needs of most principals and teachers and is little used by them. Meanwhile, teachers and students spent considerable time taking teacher-made curriculum embedded tests — tests which reflect the instructional programs and which serve the classroom decisionmaking needs of teachers, but which have little impact in the policy arena. In other words, both teachers and policymakers devote considerable attention and resources to testing, but view each others' efforts as invalid for their purposes.

While several reasons for this mutual rejection have been described above, the fact remains that both teachers, principals, district administrators, and other policymakers require information about the same phenomena: the academic progress of students and the extent to which students are achieving the skills which teachers and schools intend to teach. And while the information needs of administrators and policymakers may differ from those of teachers and principals — i.e., needs for generalizable, comparative information vs. ideographic information which is sensitive to local context — both share the need for validity. Yet the validity of achievement tests are valid measures of school progress and of accountability only under very special conditions: where their content matches the specific instructional intentions of schools. Ultimately, then, the information needs of teachers and policy-makers may be very similar, although their roles and respective responsibility implies considerably different levels of specificity and periodicity in assessment.

Given this similarity in essential information needs, it should be possible to design, in place of overlapping requirements and duplicative efforts, multipurpose testing systems which can simultaneously serve the needs of both policymakers and local educators. Such testing systems might provide very detailed and frequent information at the classroom level and for the local school site, but be combined and aggregated for decisionmaking purposes at other levels. For example, a test might provide a teacher with detailed diagnostic information about a student's strengths and weaknesses in reading objectives targetted for classroom instruction; the results of that test could also be aggregated by instructional group or class for classroom decisionmaking, be combined over time for the class and grade for school-level planning and then summarized for district-level purposes. Given the common accessibility of micro-computers in schools and their capacities for scoring, storage, retrieval, analysis, reporting, and transmission, the technology for implementing such systems is available and feasible for measures which are common across classrooms and schools. Calibrated item banks, anchor items, and meta-analysis techniques may someday permit more pecularistic data to be aggregated for decisionmaking at the individual, class, school, district, state and federal levels. These possibilities deserve exploration now, toward a more rational, integrated assessment system in the future.

This is a long-range agenda. In the short-run however, school districts can make a start in making external tests more relevant for school- and classroom-level planning and/or in building internal (classroom) tests that are useful in schoolwide and districtwide planning and decision-making. The final chapter of this monograph describes some productive models that districts can follow toward these ends.

CHAPTER 8

DIRECTIONS FOR POLICY AND PRACTICE AT THE LOCAL LEVEL: LINKING TESTING WITH INSTRUCTIONAL PLANNING AND IMPROVEMENT

In explaining the policy orientation underlying the Test Use in Schools Study, Chapter 1 listed several questions that are extremely common among and urgent for policy makers in local school districts. To restate those concerns here: many school districts are expanding their own testing programs. From district to district, however, teachers may differ in their willingness to administer these tests and to utilize their results. Under what conditions, then, are district tests most likely to be administered and used? What questions should tests have in order to make them attractive and useful from teachers' points of view. How can district testing be effectively integrated with other assessment activities?

This chapter suggests answers to these questions as it addresses a somewhat broader one: How can districts and schools make more effective use of test results in instructional planning and improvement? The models and guidelines presented below are derived not only from the general survey and detailed fieldwork findings of the Test Use in Schools Study, but also from the on-site case studies of a complementary CSE project which examined district organization and management strategies for promoting test use (Bank & Williams, 1981a, 1981b, 1983).

These field studies demonstrate ways in which the utility of both the external and internal tiers of assessment (as described at the outset of Chapter 7) can be enhanced in local decision-making and in planning for instructional improvement. There are, the data suggest, two approaches that districts can follow to accomplish this goal. *One appproach is to build from the inside out*: to construct district tests that have the characteristics of internal assessment tools — the validity for local curricula, suitability for routine classroom purposes, and immediate availability that appeal to teachers — and at the same time provide consistent, reliable data that can be aggregated in ways useful for school and district decisionmaking. *The second appproach is to build from the outside in*: to analyze information from externally mandated measures currently given in the district and deliver it to schools at times and in formats that maximize its utility in planning for curricular and instructional improvement.

These approaches are not mutually exclusive; both can be followed simultaneously. But the effectiveness of either depends upon more than the proper handling of testing and test scores. It also depends upon district systems that structure and support the use of testing information in an on-going planning process — systems of a type that are not widely present in most districts today.

On the whole, as has been shown, most districts do not routinely return test results to schools in ways that facilitate their use in decisionmaking. Administrators review scores for the faculty in most schools, but rarely on a periodic basis as part of routine procedures. Follow-up to assure that teachers are giving attention to the content area, skills, etc., that test scores indicate need emphasis is rarely routine, either. (See Table 20, page 49.) Survey data show that the majority of teachers are instructed in how to administer tests and that they are informed about test results. Yet it appears that few receive training in how to link teaching and testing or in how to use test results in improving instruction. (See Chapter 4, Table 23, page 54.) These are only some very general indicators that not many districts are closing the testing-instruction loop with systematic planning mechanisms. They are supported, however, by fieldwork from both the Test Use Study and the other CSE project mentioned above. Furthermore, even though efforts of the kinds investigated in this study are only the most elemental in a district testing-instructional decisionmaking linkage system, they can make a difference in how teachers view and use testing. Analysis of survey data show that where there is more support by district and school leaders for the use of test results in planning, and where there is more staff development in assessment, teachers have a significantly more positive view of testing and its uses, and they also tend to treat the results of district-objectives-based, standardized, and even minimum-competency tests as more important in instructional decisionmaking. (Review Table 31, page 84.) With this in mind, discussion turns to some ways that districts can create successful links between testing and planning for instructional improvement in their schools.

Building Links From the Inside Out

Districts that follow this approach build outward from classroom assessment needs to those of the school and district. They also build from what should be taught to what should be tested. First they construct district curricula, then district tests to match.

Two of the districts studied closely by CSE's projects were especially successful in taking this approach. Their slightly different testing-instruction linkage systems are useful models for others.

The Central City Model*

Located in the rural midwest, Central City School District serves about 5,000 students in seven elementary schools, three junior highs, and a high school. It has a long history of innovation and commitment to curriculum development. It also has a group of teachers who pioneered use of the high school's main-frame computers (originally purchased and used for computer-assisted instruction) in the scoring and analysis of teacher-made tests. These factors, and an energetic leader, joined in the creation of Central City's system for linking test information with instructional planning.

The test information. Each summer in recent years, the district has sponsored curriculum development projects. But while the district initiated, compensated, and guided, it was teachers who did the work. Several representatives from the faculties of each school were selected by their peers to participate.

Efforts began with the construction of an elementary-grade media (or library) skills module and continued through the development of complete mathematics and social science curricula for the elementary grades. Later, the mathematics curriculum was extended through grade 8 and work began on a reading program. In each case, development was done unit by unit in several stages. First, teachers decided on instructional objectives and selected and/or wrote materials and learning activities for achieving them. Then, pre- and post-tests referenced to the objectives of each unit were designed and "mastery levels" for each objective were specified. Units and accompanying tests were piloted the next year; objectives, materials, and test items were revised in light of teachers' criticisms and suggestions. Further revisions incorporating teachers' feedback were made after the units went into general use in schools across the district.

---

*The district names used here are pseudonyms. Any resemblance betwen these names and those of actual districts and communities is unintended.

Testing materials were designed such that all the unit tests could be scored and analyzed by computer and returned to the teachers in a day or two. Results came in the form of a set of easy-to-read sheets, one for each student. The sheet listed each objective covered on the test, the number of items that measured the particular objective, the number of these items the student had correct and incorrect, and whether the number correct equaled "mastery." At the top of each sheet appeared a paragraph that described the types of errors the student had made and summarized the types of difficulties the student seemed to be having with the skills or content covered.

In mathematics, the district had selected a sample of items from the unit tests and combined these to create mid-year and end-of-the-year summary measures given to students in all schools. Teachers received summary sheets of the type described above for these tests, too. (The district was considering developing similar tests in other subject areas once the process of curriculum and test-item revision was considered complete.)

All this applies to the lower grades, but similar developments had begun in the high school mathematics department. These were initiated by the teachers, who had worked toward common curricula and devising computer-scored tests for various courses. In line with a general district attitude, other departments were encouraged, but not required, to follow this example.

The end results of the district-wide effort were several: (1) curricula that were consistent across the district, that teachers were invested in, and that teachers actually used; (2) a system of tests that fit the curricula and provided timely information in a form appropriate for a variety of routine instructional decisions; and (3) a body of test information that was valid and consistent from classroom to classroom and could thus be aggregated and compared in school and district planning.

The structure of school decisionmaking. Within the schools, these test data came into play in two main ways. First, they were routinely used by teams of teachers in regular "unit" meetings. Elementary-school "units" included several teachers (one of whom was chosen as unit leader), a cluster of students across two or three grades, and occasionally an instructional aide. Students were often divided among unit teachers in different groupings for different subjects based on their current level of achievement and rate of learning. (Some schools, however, tended to use the self-contained classroom approach in some grades.)

Unit teams met at least weekly during release time at the end of an abbreviated school day. At the beginning of the year, they discussed students' placement and planned instructional emphases and pacing. Later on, they routinely examined students' progress, reviewed their placements, re-evaluated and altered their teaching, and discussed individual learner's problems and how best to address them. Data from district tests, as well as other available information, were routinely examined as these matters were considered. Unit meetings, then, were the primary setting for linking test data with instructional decisionmaking. (Where classrooms were self-contained, teachers reported using the district tests individually, as well as in unit meetings. And similar procedures were followed in the junior high and high school math departments.)

A second use of district test data occurred periodically as principals established school goals and agendas for school in-service activities.

District support systems. The linkage effort described above was supported by the Central School District in a number of ways.

First, district leaders initiated and provided resources for the curriculum-and-test development. They also gave release time for weekly unit meetings in which the test data were used for instructional planning.

Second, district administrative leaders provided staff development in curriculum writing and test development. Originally, these weekly, semester-long, courses were led by professors from a state university. Later, however, the district encouraged teachers to take over the classes: to revise them, make them more practical and relevant for district staff, and then to teach them. Credit on the district's pay scale was given for participation in these classes.

Third, district administrator guaranteed on-going technical assistance by maintaining close contact with the nearby Intermediate Educational Agency (IEA). IEA help was routinely sought on problems in test development and on scoring-and-analysis issues. The IEA also provided some staff development in instruction.

Fourth, the district maintained media centers staffed by instructional specialists in each school. Specialists helped unit teams and individual teachers locate supplementary teaching materials to address learners' needs. They also offered training in such areas as instructional diagnosis and prescription.

Fifth, a district administrator worked with teacher committees in piloting curriculum units and tests, eliciting teachers' critiques, and revising objectives, materials, and test items.

It was this same administrator who encouraged continuing and broadening the use of the computer-scoring-and-test-analysis process.

The Shelter Grove Model

The Shelter Grove Unified School District is located in the southwestern region of the country. Until three years ago, Shelter Grove was an elementary school district. The recent merger with a local secondary school district brought Shelter Grove's enrollment to about 5,700. These students are distributed through four elementary schools, two middle schools (grades 6-8), and a four-year high school.

Shelter Grove's system for linking testing with instruction is similar to Central City's in several ways. Yet it is different enough to be worth description as a second "inside-out" model.

The test information. Like Central City, Shelter Grove administers tests of several types. But those that have the greatest power to influence instruction in Shelter Grove schools are those developed by the district and referenced to its continua (or sequences) of instructional objectives in reading, mathematics and writing (composition).

Shelter Grove initially contracted with a commercial firm which promised to write test items for district-selected objectives and to provide computer printouts of scores. Introduced in the early 1970's, these tests failed to win teacher support. Teachers complained that the tests were not coordinated with anything that was taught. They also found that they did not know what to do with the results.

Teacher committees were appointed to try to revise test items. They responded to the perceived need to align the coordinating tests with their curriculum by beginning to work on a district-level continuum of objectives. From then on Shelter Grove's experience paralleled the more recent history of Central City. By the late 1970's, teacher committees had devised continua of objectives and accompanying criterion-referenced tests for reading and math, as well as similar tests for language arts. More recently, a district writing continuum was established.

Unlike the Central City materials, Shelter Grove's tests do not serve as unit pre-tests or post-tests. And except in written composition, district objectives are not accompanied by district-designed materials or recommended learning activities. Rather, the continua are aligned with the commercial reading and math text series used districtwide.

The district tests were routinely administered to students by classroom teachers on two or three occasions between October and February. Scores were aggregated by the district's Testing Coordinator for individual students, instructional groups, entire classes, and the school.

These profiles were sent to the schools in time for planning days that occurred regularly at several points through the year.

In addition, proficiency tests composed of various segments of the district's criterion-referenced tests were administered to children in grades 4, 5 and 6 each year in April and May in accordance with state requirements.

The structure of school decisionmaking. District tests were routinely used in each elementary and middle school during planning days that occurred at several points in the school year. (The system had not yet been introduced in the district's high school.) Two of these days were in June. On the first, the program of the school was routinely evaluated by the entire school staff looking at the group, classroom, and total school scores. These sessions functioned as a needs assessment for the next school year. On the second June planning day, individual teachers placed students in appropriate learning groups for the coming year using the test-result profiles on each student.

In September of each year, test information was updated; information on students new to the district was added. In October, teachers met with their principals to set learning goals — benchmarks on the continuum that, based upon past performance profiles, they expected the children in each instructional group to meet.

A mid-year evaluation took place each February. Summary reports on current-year testing were run, distributed, and examined. Principals met with teachers, as well as with the Superintendent and Assistant Superintendent for Instruction, to discuss students' progress. Plans for modifying the instructional program were made at this time. Then, in June, the cycle began anew with reference to the again-updated test-score profiles.

Individual teachers also used criterion-referenced test information in reporting to parents each October and again each spring. Report cards listed continuum skills on one side and noted students' progress toward each objective. And each May, letters were sent to the parents of children who were two grade levels behind expected performance; special conferences with these parents were also arranged.

District support systems. As was the case in Central City, a number of district activities and programs helped to sustain the linking of test data with instructional planning in Shelter Grove. In addition to the district's leadership and resources in developing the instructional-objectives continuua and criterion-referenced tests, these included the following.

First, the district maintained a Professional Development Program (PDP) that provided teachers with the skills necessary to act upon the test results. Coordinated by a full-time specialist, the PDP had evolved over time based upon the Madeline Hunter orientation to teaching. Level One activities (for all new teachers, aides, and substitutes) dealt with such basic teaching skills as understanding goals and objectives, motivation and reinforcement, and task analysis and diagnosis. Level Two activities (which were not required but encouraged, and which many teachers joined) extended those of Level One with emphasis on individualizing instruction. Strategies for meeting affective needs, using inquiry skills, and teaching specific curriculum content were also covered. The program required teachers to apply PDP skills in their own classrooms, with supervision and feedback from the PDP coordinator. (Prior to the general implementation of this PDP program, all principals had been required to take the Level One course plus courses in clinical teacher supervision.)

Second, learning specialists conducted demonstration lessons, recommended materials, conducted diagnoses of new students, and assisted teachers in planning and placement when new criterion-referenced test scores arrived in the schools. The learning specialists were considered master teachers, and regularly played an important role in helping teachers use test information. They also explained changes in the continuum or changes in district policy to the faculty. With the PDP, learning specialists were perceived as critical supports to the district's linkage effort.

Third, a Testing Advisory Committee composed of a principal and several teachers continually updated and improved the district's tests in light of teacher criticisms. This group also handled whatever administrative and technical problems arose in testing, scoring, and reporting results.

Fourth, *ad hoc* continuum revision committees made up of teachers and learning specialists were paid during the summer to revise sections of the continua as seemed appropriate.

In addition to these formal organizational features, a variety of other networking activities (e.g., principal observations, learning specialists' visits to classrooms, monthly meetings of a district communications council) helped district personnel work closely together in maintaining links between test data and instructional planning in the Shelter Grove schools.

Guidelines

The experiences of Central City and Shelter Grove, especially in contrast to those of two other districts with similar but less successful linkage systems (to be mentioned below), suggest a number of guidelines for other districts to follow in linking testing with instruction from the inside out.

1. *Build curriculum and assessment measures together "in-house."*

Administrators and teaching staff in both districts believed very strongly in the district development process. They felt that it helped assure teacher "ownership" and confidence in both curricula and tests; ownership and confidence, in turn, seemed to be important prerequisites for teacher use. Shelter Grove's unhappy experience with tests built outside the district, even when they were developed to district specifications, supports this wisdom.

2. *Assure a close fit between test items and curricular objectives and materials.*

This can best be done by designing curriculum first and then the tests, as was done in Central City and, ultimately, in Shelter Grove as well.

Teachers are inclined to see district objectives-based or criterion-referenced tests as a burdensome irrelevancy if this condition is not met. New Branford, an urban district with 30,000 enrollment in the northeastern United States, attempted to devise criterion-referenced tests keyed to its district reading and math objectives. But when Test Use in Schools researchers visited New Branford schools, they found that few teachers used these tests. Continuum objectives were intended to fit with all of the five or six math and reading series used across the district. In fact, according to teachers, they fit well with none of them. Thus, teachers continued to use the tests included with these commercial series to get the information on achievement they needed — and they also had to give district tests to comply with district requirements. But information from the latter was rarely consulted, and teachers resented the mandate to give them. For similar reasons, Central City teachers neglected their district's objectives-based reading tests, although they were generally enthusiastic about those in the other subjects, developed years earlier with little teacher participation and without accompanying curriculum materials. Teachers complained that the reading tests were no longer valid for the two basal reading series used in Central City.

3. *Strive for maximum teacher involvement.*

To help build curriculum and tests that teachers own and use, teachers' participation in the development process must be more than nominal. Both Shelter Grove and Central City included many teachers on their development committees; these teachers did the real work of constructing the curricula (or continua) and the test items. Mechanisms were provided that allowed all district teachers to offer feedback on a regular basis. Their criticisms were taken seriously in the revision process.

In contrast, New Branford (mentioned just above) and Metro District (another urban district studied by the CSE Test Use Project) had only a small number of teachers on district advisory committees as they constructed continua of objectives and accompanying tests. These teachers did not participate in the actual development process; their presence was not visible to district faculty; they had little impact on the tests that evolved. And in neither district did teachers feel the objectives or tests were completely suitable. New Branford teachers' response has been described. Teachers' response to Metro District's tests was quite mixed.

4. *Construct tests that cover the entire range of skills in the curriculum and/or continuum of objectives.*

The district tests of Central City and Shelter Grove included items that assessed students' performance on skills and content from the most elemental to the most advanced in the subject areas tested. Metro District (enrollment over 100,000), in contrast, purchased tests for each grade level in reading, math, and language arts that covered only the simplest skills to be taught. In the economically disadvantaged neighborhoods where more students had trouble with these skills, test results did help teachers identify the skills which individuals and class groups needed remediation. But in these schools, the tests also functioned to push the actual curriculum in the direction of the most elemental skills. Teachers and principals wanted students (and their schools) to do well on the tests each spring. Thus, they spent much time drilling and re-drilling children on the elemental skills tested. Simultaneously, they gave shorter shrift in their teaching to other skills specified for the grade level, which were not included on the test. Elsewhere in the district, where students routinely obtained 90 percent to 100 percent correct on these same tests, they yielded little diagnostic or placement information for teachers.

One moral of these contrasting stories, then, is test what you want teachers to teach, because teachers will place their teaching emphasis on what you test.

Several other "do's" and "don'ts" can be abstracted from the Central City, Shelter Grove, and similar but less successful models. These, however, are equally pertinent to the "outside-in" linkage approach discussed next. Thus, they will be omitted here and mentioned in the concluding summary.

Building Links From the Outside In

Districts that follow this approach adapt information from externally mandated tests to suit the district's and/or schools' planning needs. In so doing, they support school-level planning structures and procedures, just as districts taking the inside-out path do.

The testing-instruction linkage systems of two districts that followed the outside in approach are described below. They provide very different, but equally instructive models.

The St. John Model

The St. John School District covers a wide geographic area of suburban and semi-rural municipalities in a Western state. Its 72 schools serve between 40 and 50 thousand students in grades K-12.

Linking testing with instructional planning began in St. John during the mid-1970's when the state legislature enacted a program intended to stimulate local planning for school improvement. Participation in the program was voluntary, but over the years most of St. John's elementary schools, along with two of its junior high schools and one high school, elected to participate. The district encouraged this involvement; in turn, the schools' participation stimulated district efforts to provide test data for use in local site planning.

The test information. Long before the advent of the state-sponsored school improvement program, St. John School District had required administration of the Iowa Test of Basic Skills. Students were tested each January in grades 2-6. The purposes this information had served previously are not germane here. But once numerous St. John schools joined the state program, test data became especially important for them. Guidelines for the state school-improvement planning process required that in establishing improvement plans schools specify: (1) the "existing level of performance" in a particular area, (2) the "needed program changes or additions," (3) improvement objectives, and (4) activities to measure these objectives. Major activities to be undertaken in pursuit of each objective also had to be described, along with budgets and other improvement program features. But the four requirements enumerated here were those that called for "hard data" such as test results.

It seemed reasonable to use ITBS results in developing these improvement plans, yet district administrators realized that these results came back from the test publisher in a form that was cumbersome.

Computer printouts presented the results for each sub-test area for each grade for each year on a separate page. Principals and teachers found these reports complicated as well as overwhelming in volume. Consequently, the district undertook development of what it now calls the Academic Performance Profile (APP).

The APP gave each district elementary school an annual overview of its ITBS test results for all years and all grades for a particular subtest (e.g., reading comprehension, math concepts, etc.) on a single page. This reduced fifty pages of computer printout to approximately six, ordinary 8 by 11 inch pages.

In addition, the APP simplified the format in which the information appeared. Simple graphs were devised to visually display : (1) the scores of student groups as they moved through the grades (1982 first graders as second graders in 1983, etc.); (2) the performance at various grade levels in various years (the fourth grade in 1981, 1982, 1983, etc.); and (3) the gains (indicated in terms of grade-level growth) realized from one year to the next for the various grade levels (the gains made by the 1982 second-grade group as third graders in 1983). Two simple tables on each page (that is, for each sub-test) supplemented the three-line graphs.

Since the state program guidelines also called for annual needs assessment, the St. John District created survey questionnaires for staff, parents, and students. These solicited respondents' perceptions of: (1) the effectiveness of schools' various programs; and (2) how much attention should be given to improvement in each program area. Each school could add up to 20 questions to the set used in common across the district. Surveys were administered annually in the spring of each year. The district's evaluation office tabulated survey results for each school and returned them in a concise form.

The structure of school decisionmaking. The state's school improvement program mandated the creation of a School Planning Council (SPC) in each participating school. Guidelines directed that the SPC membership include the principal and elected representatives of the teachers, of other school staff, of parents and other community members, and (at the secondary level) of the student body. This group was assigned central responsibility for establishing needs, goals, and activities for school improvement, as well as for budgeting the state funds provided to the school for improvement activities.

St. John's district evaluation specialists, however, elaborated on these state requirements. They urged their schools to also create "component committees," smaller groups (including SPC members and others) who were charged with planning for improvement in particular areas — in each subject area, in school environment, in human relation, in staff development, etc.

Component committees reviewed the ITBS/APP summary forms, survey results, and other information. They specified and documented needs, set objectives, and developed school and classroom activities to realize them. They also stated how achievement of the objectives would be evaluated and proposed a budget suitable for their plan. In a next step, various component committees presented their particular plans to the School Planning Council. The SPC accepted or suggested changes in each improvement-plan component and made decisions regarding final allocation of state program dollars among the various components. The SPC also monitored implementation of the plan through the coming school year.

While plans were routinely developed for a three-year period, revisions were made each spring based on information gathered during the current school year. Thus, school improvement planning was an annual process centered in the spring, but implementation of plans and SPC monitoring occurred continuously during each school year.

Interviews with participants and observation of planning meetings indicated that test data (and survey results) were used in deciding upon and substantiating needs, specifying objectives, evaluating implementation, and revising the plans. SPC members also routinely referred to this information in making and justifying budgetary decisions.

District support systems. The St. John School District supported its testing-instruction linkage system in many of the same ways that Shelter Grove and Central City supported their quite different models.

First, staff development in the organization and process of planning, including the use of the APP test summaries, was conducted for 600 district personnel during their first year in the state program. Others received this introductory training as they entered the program. Furthermore, teachers, principals, and parents agreed that the regular availability of the districts' two evaluation specialists was a key to the program's maintenance. They routinely provided staff development and answered *ad hoc* questions regarding planning and test-data use.

Second, St. John maintained a comprehensive staff-development program in instructional techniques, which everyone agreed was a major factor in facilitating the realization of school plans.

The Bayview Model

Bayview is a community of 100,000, and is located about 50 miles from a major Western metropolitan area. The Bayview Unified School

District's sixteen elementary schools, four junior highs, and three senior highs enroll 14,000 students.

Bayview's six-year-old effort at testing-instructional linkage was more diffuse than that in most of the other school districts visited by CSE researchers. Interest in testing and evaluation was relatively new, and many in the district were as yet skeptical of their value. Nonetheless, the need to comply with externally mandated testing programs stimulated a small group of district administrators to try to make greater local use of the test scores they yielded. Only one of these uses will be discussed here. It offers an example of "outside in" testing-instruction linkage that is quite different from the St. John School District's model.

The test information. Three different achievement testing programs figured in the Bayview linkage system are described here. The first of these was the State Assessment Program (SAP). This half-hour test was administered each spring to students in grades 3, 6, and 10 in accord with state requirements. The test was devised by the state and referenced to objectives common to many state-approved text series. Items were matrix sampled; not every student was asked to respond to identical questions. Thus, data for individual students were not reported. Results focused on grade-level and school patterns.

A second test used by Bayview was the norm-referenced, standardized Comprehensive Tests of Basic Skills (CTBS). The district had just begun to require this test in all schools for grades 1-9 when CSE fieldwork was conducted. Formerly, it had been given only in schools with Title I (now Chapter 1) compensatory education programs.

The district's proficiency (or minimum competency) testing program was also used in testing-instruction linkage. Forms for grades 5, 9, 10, and 11 had been developed with the help of consultants to meet the state's mandate. These measures covered reading, writing, and mathematics skills deemed essential for "life coping." The current forms of the test were introduced in 1978.

The decision-making structure. The data from these three tests were brought to bear on instructional planning in several ways by Bayview district leaders. Chiefly, however, they had begun to use the three test programs mentioned above as content for staff development course work in task analysis and diagnostic-prescriptive teaching.

District leaders had won grant funds from the state to create a Professional Development Center (PDC). The primary focus of the PDC's program was the continuing development of effective teaching strategies. A Teacher Center funded by a federal grant augmented the PDC. Curriculum development and the translation of educational research for practical, instructional applications were the central thrusts of the

Teacher Center's program. The very presence of these two centers testified to Bayview's emphasis on teaching-effectiveness skills. In addition, principals were required to attend workshops dealing with supervision, and these focused on the elements of effective teaching.

It was in the context of increasing external test mandates and the emphasis on staff development that Bayview's linkage system began to take shape. From the perspective of District leaders, Bayview teachers and principals were not facing the issues raised by the District's relatively poor performance on the external measures. In response, said the Director of Staff Development:

> We [at the central office] tried to model a problem-solving way of looking at it so principals could do similarly in their schools. The Director of Instruction worked with principals in the way he wanted them to work with teachers. Also, we asked teachers if they were addressing areas of the test. They said they were. When we observed, we found teachers had difficulty defining the skills to be taught as well as diagnosing for these skills. As a result, we built task analysis cycles into our Professional Development Center programs focusing on the low scoring skill areas identified by the State Assessment Program.

The district's cadre of leaders began by training principals to examine SAP (and later the other tests mentioned earlier) to see what specific skills they assessed. Once these were identified, the next step was for principals and faculties to examine their school's curricula in order to determine whether these skills were being taught and if so at what grades and with what emphasis. Staff development provided principals, and later teachers, with the information and techniques they needed to do this.

This was taking place with varying degrees of thoroughness in different Bayview schools when CSE staff members visited the district. At the same time, areas of curricular and instructional weakness districtwide had been identified by district administrators. These areas were then targeted for sessions on diagnostic-prescriptive teaching and other instructional skills.

Analysis of test results also suggested areas for emphasis in the development of continua. Citing the impact of proficiency-test skill and score analysis, for example, the Bayview Coordinator of Curriculum said:

> The proficiency exam has helped the district focus on curriculum. . . [We learned that] in math we teach computation but the test tests applications through story problems.

Thus, in the Bayview Unified School District, task analysis of tested skills served as the basis for a comprehensive examination of the district's curricula and suggested areas of curricular weakness. Simultaneously, analysis of test results led to the identification of teaching weaknesses. Links between testing and instruction were generated through the development of district-wide objectives and in Professional Development Center and Teacher Center programs.

Guidelines

The St. John and Bayview districts had put in place very different kinds of systems for linking the results of externally mandated testing with instructional planning in their schools. Nevertheless, it is possible to abstract a number of guidelines from their "outside-in" models. Other districts would be well advised to bear these in mind should they chose to follow the outside-in approach.

1. *Make test-score data comprehensible for teachers and principals.*

Providing test results in a format that facilitates their use is obviously a key to testing-instruction linkage. That professional educators working in the schools can be bewildered and intimidated by reports of scores from externally mandated measures was clear in Test Use in Schools Study fieldwork (cited early on in this paper). It was equally apparent in the early experiences of district administrators in both Bayview and St. John. The latter addressed this problem by translating the scores into succinct, easy-to-read, and relevant tables and graphs. Bayview dealt with it by teaching principals and teachers to dissect the tests and test results.

2. *Train teachers and principals to use test scores as diagnostic tools.*

As noted earlier, the results of externally mandated tests are commonly used in a brief and casual way to get a general comparative reading on group performance. The essence of their use in the St. John and Bayview systems was diagnostic. They played a role in identifying patterns of strength and weakness in particular content areas and skills. They served to stimulate questions such as "Why are we scoring as we are scoring in this curriculum area?" and "How can we improve?" Diagnostic uses are not routine in most schools. Simply presenting test scores in clear, readable format does not mean that diagnosis of curricular strengths and weaknesses will occur. Teachers need instruction and practice in analyzing the different factors that underlie test performance. They need instruction and help in abstracting meaning from scores. Survey findings suggest that most districts do not provide this. In different ways, both St. John and Bayview did.

3. *Expect that results of externally mandated tests will serve as only one source of information in planning and decision-making.*

Wisely, neither Bayview's cadre of leaders nor St. John's district evaluation specialists tried to make test results the sole basis for educational decisions. Human values and priorities do and should influence decisions about what objectives to pursue in school improvement or to build into district continua. The day-to-day experiences with students that teachers and principals rely upon so heavily are very relevant in making instructional decisions. These factors were routinely accepted, along with test data, as bases for decision-making by St. John administrators as they assisted School Planning Councils and reviewed their plans. Bayview's Coordinator of Staff Development, too recognized that test data needed to be examined in light of other factors as he explained, "When we see through our task analysis and curriculum review what we are and are not teaching, the next step is to ask, "Do we or don't we want to teach this? How important is it for our students?"

Data from externally mandated tests can serve to identify problems, to support or disconfirm experience-based judgments, and to stimulate questions. It can be used to justify or rationalize decisions that have already been made. But as the separate experiences of St. John (recall their needs assessment questionnaires) and Bayview (recall their juxtaposition of multiple measures to district curricula) indicate, test data in themselves are only one important source of information for educational planning.

Summary and Conclusions

CSE's national survey and its fieldwork in two research projects suggest that both testing that is internal to the school and that which is externally mandated can be used more fully in systematic educational decisionmaking. Districts can build a curriculum and tests that can serve teachers' routine classroom needs *and simultaneously* provide consistent, reliable, and valid data for school and district planning. Districts can also capitalize upon data from externally mandated testing by adapting it to local needs. No single approach or model will be appropriate to every setting. But whether a district chooses to pursue linkage from the inside out or from the outside in, there are several factors that seem necessary for success.

One of these is district leadership. In each district studied by CSE, there was an individual or a small group in the district office — idea champions and supporters — who were vitally interested in using test data in instructional planning and decisionmaking. CSE's national test use survey substantiates that such leaders make a difference in school-level uses of test information.

A second element in district success is an organizational arrangement — a setting and set of procedures — for decision-making. In Central City schools there were the weekly meetings of unit teams; in St. John, regular sessions of the School Planning Councils. Shelter Grove held its principal-teacher planning days in June, October, and February each year. In Bayview, the locus of linkage was staff development workshops, continuum-building committees, and regular school faculty meetings. These organizational arrangements motivated and structured the use of test results by creating (1) real needs for information, and (2) procedures by which the implications of test-score patterns could be discussed and acted upon. None of the districts with successful linkage systems simply offered schools test data and left their use to chance.

Third, each of the districts managed testing and/or test results such that they increased the marginal utility of test information for teachers and principals. Teachers routinely receive data on student achievement as they watch their students in class, review their assignments, and grade classroom tests. These data are immediate, rich, and compelling. So too is the information principals regularly gather as they talk with staff and visit their classrooms. To be as useful and as compelling, external test information must add "something new" to what teachers and principals already know. Each of the four models described above did this. Central City's computer-scoring-and-analysis system for unit tests summarized individual students' mastery of objectives, as well as their errors and weaknesses. Shelter Grove compiled data on the progress of individuals and instructional groupings toward benchmark goals. St. John's Academic Performance Profiles charted year-to-year trends and annual gains. Bayview's task analysis projects, based on tested skills and test scores, helped to reveal why and how students' performance came to be as it was. In each case, test data was configured in ways that told teachers and principals something more than "your students are doing well in this and not so well in that" — which is information teachers and principals typically feel they already have.

A fourth and final element in successful district linkage is the maintenance of on-going resource and support systems. In the districts studied, these centered in the area of staff development: training in test development and use, training in how to realize instructional goals derived from test information, or both. Frequently, too, instructional support staff — learning specialists, media specialists, evaluation specialists — were routinely available to provide help and answer questions. Support also took the form of adaptability and flexibility on the part of district administrators. Clear channels were open for Central City and Shelter Grove teachers to participate in the development of, and to criticize the quality of district curriculum and tests. St. John's evaluation specialists revised district needs-assessment surveys in light of teachers' feedback; local schools could add survey items suitable to their particular concerns. Bayview district leaders showed patience and understanding in encouraging principals and teachers to take a "problem-solving approach" to low test scores. And of course, each district supported its testing instructional linkage system with release time and other resources.

The models and guidelines suggested here will not answer all the questions and concerns school districts will encounter as they work systematically to link testing and instruction in an on-going process of school renewal. But they do indicate productive paths toward the more efficient use of testing and the improvement of educational planning in American schools.

REFERENCES

Airasian, P.W. (1979). *The effects of standardized testing and test information on teachers' perceptions and practices.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Baker, E.L. (1978). *Is something better than nothing? Metaphysical test design.* Paper presented at the 1978 CSE Measurement and Methodology Conference, Los Angeles.

Bank, A., & Williams, R.C. (1981a). *Evaluation design project: Organizational study* (Annual Report, 1980-1981). Los Angeles: UCLA Center for the Study of Evaluation.

Bank, A., & Williams, R.C. (1981b). *Evaluation in school districts: Organizational perspectives.* CSE Monograph Number 10. Los Angeles: UCLA Center for the Study of Evaluation, 1981.

Bank, A., & Williams, R.C. (1983). *Assessing the costs and impacts of managing T/E/I Systems: A collection of nine papers.* Los Angeles: UCLA Center for the Study of Evaluation.

Berman, P., & McLaughlin, M.W. (1977). *Federal programs supporting educational change, Vol. III: Implementing and sustaining innovations.* (R-1589/8-HEW). Santa Monica, CA: Rand Corporation.

Boyd, J., Jacobsen, K., McKenna, B.H., Stake, R.E., & Yashinsky, J. (1975). *A study of testing practices in the Royal Oak (Michigan) Public Schools.* Royal Oak Michigan: Royal Oak Public Schools.

Burden, J.L. (1982). Teacher certification. In H.E. Mitzel (Ed.), *Encyclopedia of educational research* (5th ed.). New York: The Free Press.

Burry, J., Catterall, J., Choppin, B., & Dorr-Bremme, D. (1982). *Testing in the nation's schools and districts: How much? What kinds? To what ends? At what costs?* CSE Report No. 194. Los Angeles: UCLA Center for the Study of Evaluation.

Center for the Study of Evaluation (1979). *CSE criterion-referenced test handbook.* Los Angeles: UCLA Center for the Study of Evaluation.

Choppin, B., Dorr-Bremme, D.W., & Burry, J. (1981) *Test use project annual report.* Los Angeles: UCLA Center for the Study of Evaluation.

Cicourel, A.V. (1974). *Cognitive sociology: Language and meaning in social interaction.* New York: The Free Press.

Cicourel, A.V., & Kitsuse, J.I. (1963). *The educational decisionmakers.* Indianapolis: Bobbs-Merrill.

Coffman, W. (1983). Testing in the schools: A historical perspective. In E.L. Baker & J.L. Herman (Eds.), *Testing in the nation's schools: Collected papers* (pp. 3-27). Los Angeles: UCLA Center for the Study of Evaluation.

Cronbach, L.J. (1963). Course improvement through evaluation. *Teachers College Record, 64,* 672-683.

Dawson, J.E. (1977). Why do demonstration projects? *Anthropology and Education Quarterly, 8*(2), 95-105.

Deal, T.E. (1979). *Linkage and information use in educational organizations.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Dorr-Bremme, D.W. (1983). Assessing students: Teachers' routine practices and reasoning. *Evaluation Comment, 6*(4), 1-12.

Ebel, R.L. (1967). Improving the competence of teachers in educational measurement. In J. Flynn & H. Garber (Eds.), *Assessing behavior: Readings in educational and psychological measurement*. Reading, MA: Addison-Wesley.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(2), 15-22.

Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. Hathaway (Ed.), *New directions for testing and measurement: Testing in the schools* (pp. 29-38). San Francisco: Jossey-Bass.

Garfinkel, H. (1967) *Studies in ethnomethodology*. Englewood Cliffs, N.J.: Prentice-Hall.

Goslin, D.A. (1965). *The use of standardized ability tests in American secondary schools and their impact on students, teachers, and administrators*. New York: Russell Sage Foundation.

Goslin, D.A., Epstein, R., & Hallock, B.A. (1965). *The use of standardized tests in elementary schools*. Second Technical Report. New York: Russell Sage Foundation.

Huron Institute. (1978). *Summary of the Spring Conference of the National Consortium on Testing*. Cambridge, Mass.: Huron Institute.

Lazar-Morrison, C., Polin, L., Moy, R., & Burry, J. (1980). *A review of the literature on test use*. CSE Report No. 144. Los Angeles: UCLA Center for the Study of Evaluation.

Linn, R.L. (1983a). Curriculum validity: Convincing the courts it was taught without precluding the possibility of measuring it. In G.F. Madaus (Ed.), *The courts, validity, and minimum competency testing*. Hingham, MA: Kluwer-Nijhoff.

Linn, R.L. (1983b). Testing and instruction: Links and distinctions. *Journal of Educational Measurement, 20*, 179-189.

Madaus, G.F., & McDonough, J. (1979). Minimum competency testing: Unexamined assumptions and unexplored negative outcomes. In R. Lennon (Ed.), *Impact changes in measurement: New directions for testing and measurement, 3*(3), 1-14.

Mayo, S.T. (1959). Testing and the use of test results. *Review of Educational Research, 29*(1), 5-14.

Mehan, H., & Wood, H. (1975). *The reality of ethnomethodology*. New York: Wiley Interscience.

Meyer, J.W., & Rowan, B. (1978). The structure of educational organizations. In M.W. Meyer and associates (Eds.), *Environment and organizations*. San Francisco: Jossey-Bass.

Montjoy, R.S., & O'Toole, Jr., L.F. (1979). Toward a theory of policy implementation. *Public Administration Review*, 39(5).

Perrone, V. (1978). *Remarks to the National Conference on Achievement Testing and Basic Skills*. Paper presented at the National Conference on Achievement Testing and Basic Skills, Washington, D.C.

Purkey, S.C., & Smith, M.S. (1982). *Effective schools — A review*. Paper prepared for the national invitational conference, Research on Teaching: Implications for Practice, Airlee House, Warrenton, VA.

Resnick, L.B. (1981). Introduction: Research to inform a debate. *Phi Delta Kappan*, 62(9), 623-625.

Rudman, H., Kelly, J.L., Wanous, D.S., Mehrens, W.A., Clark, C.M., & Porter, A.C. (1980). *Integrating assessment with instruction: A review 1922-1980*. East Lansing, MI: Institute for Research on Teaching.

Salmon-Cox, L. (1981). Teachers and tests: What's really happening? *Phi Delta Kappan, 69*(9), 631-634.

Schutz, A. (1962). *Collected papers I: The problem of social reality.* The Hague: Martinus Nijhoff.

Stetz, F., & Beck, M. (1979). *Comments from the classroom: Teachers' and students' opinions of achievement tests.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Stinnett, T.M. (1969). Teacher certification. In R.L. Ebel (Ed.), *Encyclopedia of educational research* (4th Ed.). New York: Macmillan.

Tinkleman, S.N. (1966). Regents examinations in New York State after 100 years. *Proceedings of the Invitational Conference on Testing Problems.* Princeton, NJ: Educational Testing Service.

Tyler, R. (1977). What's wrong with standardized testing? *Today's Education, 66*(2), 35-38.

Weatherly, R., & Lipsky, M. (1977). Street-level bureaucrats and institutional innovation. *Harvard Educational Review, 47*(2).

Weider, D.L. (1973). *Language and social reality.* The Hague: Mouton.

Woellner, R.S. (1979). Let's use tests for teaching: Standardized test results can provide the basis for a program of instruction. *Teacher, 90*(2), 62-64, 179-181.

Wood, H. (1968). *The labelling process on a mental hospital ward.* Unpublished master's thesis, University of California, Santa Barbara.

Yeh, J.P. (1978). *Test use in schools.* Los Angeles: UCLA Center for Study of Evaluation.

Yeh, J.P. (1980). Reanalysis of data. In D.W. Dorr-Bremme (Ed.), *Test use project annual report* (Vol. II). Los Angeles: UCLA Center for the Study of Evaluation.

Yeh, J.P., Herman, J.L., & Rudner, L.M. (1981). *Teachers and testing: A survey of test use.* CSE Report No. 166. Los Angeles: UCLA Center for the Study of Evaluation.